

A>Note: this sequence may represent an allelic form or another gamma chain subclass
C:Comment: The heavy chain disease protein Wis is shown.
C:Genetics:
A:Gene: GDB:IGHG3
A:Cross-references: GDB:119338; OMIM:147120
A:Map position: 14q32.33-14q32.33
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; immunoglobulin pyroglyutamic acid
F:203-270/Domain: immunoglobulin homology <IM2>
F:1/Modified site: pyrrolidone carbonylic acid (Gln) #status experimental
F:6,140/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 46.9%; Score 1133; DB 1; Length 289;
Best Local Similarity 88.2%; Pred. No. 4.6e-62;
Matches 209; Conservative 13; Mismatches 9; Indels 6; Gaps 2;

QY 201 PCF-AEPKSCDKTHTC-----PELLGSPVFLPPPKDITLMSRTPVTCVVVDVSH 254
DB 53 PCFRCPEKSCDTPPCPCPAPELLGSPVFLPPPKDITLMSRTPVTCVVVDVSH 112

QY 255 DEPVKNVYVDGVEVNAKTKPREBOYNSTRVSVLTVLHQLWLNKGYKCAVSNKALP 314
DB 113 DEPVQKMYVDGQVNAKTKPREBOFNSTRVSVLTVLHQLWLNKGYKCAVSNKALP 172

QY 315 APIEKTISKAGQPREPOVYTLTPPSDELTKNOVSLTCLVKGYPSPDIIVEMESNGOPEN 374
DB 173 APIEKTISKAGQPREPOVYTLTPPSDELTKNOVSLTCLVKGYPSPDIIVEMESNGOPEN 232

QY 375 NYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNFSCSVMEALHNYTKSLSLSPG 431
DB 233 NYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNFSCSVMEALHNYTKSLSLSPG 269

RESULT 8
IGHU
Ig gamma-2 chain C region - human
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1981 #sequence revision 13-Jun-1983 #text_change 21-Jul-2000
C:Accession: A93906; A92809; A90752; A93132; A02148
R:Ellison, J.; Hood, L.
Proc. Natl. Acad. Sci. U.S.A. 79, 1984-1988, 1982
A:Title: Linkage and sequence homology of two human immunoglobulin gamma heavy chain con
A:Reference number: A93906; MUID:82197621; PMID:6804948
A:Accession: A93906
A:Molecule type: DNA
A:Residues: 1-326 <ELL>
A:Cross-references: GB:J00230; NID:932759; PIDN:CAB58438.1; PID:6066056
A>Note: Lys-326 is probably removed posttranslationally
R:Wang, A.C.; Tung, E.; Fudenberg, H.H.
J. Immunol. 125, 1048-1054, 1980
A:Title: The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and E
A:Reference number: A92809; MUID:81007873; PMID:6774012
A:Contents: myeloma protein T11
A:Accession: A92809
A:Molecule type: protein
A:Residues: 119, 'Q', 21-57, 'Z', 59, 'A', 61-193, 'D', 195-325 <MAN>
A>Note: Trp-156 is at or near the complement-binding site
R:Connell, G.E.; Parr, D.M.; Hofmann, T.
Can. J. Biochem. 57, 758-767, 1979
A:Title: The amino acid sequences of the three heavy chain constant region domains of a
A:Reference number: A90752; MUID:80001357; PMID:113060
A:Contents: myeloma protein Z1e
A:Accession: A90752
A:Molecule type: protein
A:Residues: 1-24, 'E', 26-57, 'EV', 60-85, 132-171, 'ZZZ', 175, 'B', 177-193, 'D', 195-196, 'Q', 198-
A>Note: this sequence has since been revised
R:Hofmann, T.; Parr, D.M.
Mol. Immunol. 16, 923-925, 1979
A:Title: A note on the amino acid sequence of residues 381-391 of human immunoglobulin g
A:Reference number: A93132; MUID:80114419; PMID:118920
A:Contents: Z1e
A:Accession: A93132
A:Molecule type: protein

A:Residues: 238-275 <HOF>
R:Hofmann, T.; Parr, D.M.
Submitted to the AIAA, March 1980
A:Reference number: A94591
A:Contents: annotation; Z1e, revisions to residues 25, 59, 60, and 264-268
A>Note: the revised sequence differs from that shown in having 60-Ala and in the amidati
ned

R:Milstein, C.; Frangione, B.
Biochem. J. 121, 217-225, 1971
A:Title: Disulfide bridges of the heavy chain of human immunoglobulin G2.
A:Reference number: A90253; MUID:72033500; PMID:4940472
A:Contents: annotation; myeloma protein Sa, disulfide bonds
A:Frangione, B.; Milstein, C.; Pink, J.R.L.
Nature 221, 145-148, 1969
A:Title: Structural studies of immunoglobulin G.
A:Reference number: A93157; MUID:69064124; PMID:5782707
A:Contents: annotation; Sa, disulfide bonds
C:Genetics:
A:Gene: GDB:IGHG2
A:Cross-references: GDB:119338; OMIM:147110
A:Map position: 14q32.33-14q32.33
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kai
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 16
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F:20-85/Domain: immunoglobulin homology <IM1>
F:133-202/Domain: immunoglobulin homology <IM2>
F:239-306/Domain: immunoglobulin homology <IM3>
F:14/Disulfide bonds: interchain (to light chain) #status experimental
F:7-83,140-200,246-304/Disulfide bonds: #status experimental
F:102,103,106,109/Disulfide bonds: interchain (to heavy chain) #status experimental
F:176/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 46.8%; Score 1130.5; DB 1; Length 326;
Best Local Similarity 82.6%; Pred. No. 7.5e-62;
Matches 214; Conservative 12; Mismatches 12; Indels 21; Gaps 3;

QY 181 TTTCTV-----LQNKVYEFK-IDYPCRPPEKSCDKTHTCELLGSPVFLPPPKP 232
DB 80 TTYCVNDHAKNSNTKDKYVERKCVCECPCPAP-----VAGPSVFLPPPKP 126

QY 233 KDTLMSRTPVTCVVVDVSHEDPEVKFMWYVDGVEVNAKTKPREBOYNSTRVSVLT 292
DB 127 KDTLMSRTPVTCVVVDVSHEDPEVOGRNWYVDGVEVNAKTKPREBOFNSTRVSVLT 186

QY 293 VLHODMLNGKEYKCAVSNKALPAPIEKTISKAGQPREPOVYTLTPPSDELTKNOVSLTC 352
DB 187 VLHODMLNGKEYKCAVSNKALPAPIEKTISKAGQPREPOVYTLTPPSDELTKNOVSLTC 246

QY 353 LVKGYPSDIAVEMESNGOPENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNFSCSV 412
DB 247 LVKGYPSDIAVEMESNGOPENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNFSCSV 306

QY 413 MHEALHNYTKSLSLSPG 431
DB 307 MHEALHNYTKSLSLSPG 325

RESULT 9
GAHU
Ig gamma-4 chain C region - human
C:Species: Homo sapiens (man)
C:Date: 02-Apr-1982 #sequence revision 02-Apr-1982 #text_change 16-Jul-1999
C:Accession: A90933; A90249; A02150
R:Ellison, J.; Buxbaum, J.; Hood, L.
DNA 1, 11-18, 1981
A:Title: Nucleotide sequence of a human immunoglobulin C-gamma4 gene.
A:Reference number: A90933; MUID:83157104; PMID:6299662
A:Accession: A90933
A:Molecule type: DNA
A:Residues: 1-327 <ELL>
A>Note: the sequence was determined from the germline gene
R:Pink, J.R.L.; Buttery, S.H.; De Vries, G.M.; Milstein, C.

A:Accession: A32722
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 26-426, 428-458 <CAM>
R:Carri, S.A.; Hemling, M.E.; Polna-Wasserman, G.; Sweet, R.W.; Anumula, K.; Barr, J.R.; J. Biol. Chem. 264, 21286-21295, 1989
A:Title: Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor
A:Reference number: A34194; MUID:90078232; PMID:2592374
A:Contents: disulfide bonds; carbohydrate-binding sites
A:Accession: A34194
A:Molecule type: protein
A:Residues: 26-394 <CAR>
R:Lederman, S.; Demartino, J.A.; Daugherty, B.L.; Foeldvari, I.; Yellin, M.J.; Cleary, A. Mol. Immunol. 28, 1171-1181, 1991
A:Title: A single amino acid substitution in a common African allele of the CD4 molecule
A:Reference number: A53287; MUID:92072595; PMID:1961196
A:Accession: A53287
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 250-264, 'W', 266-280 <LED>
A>Note: sequence extracted from NCBI backbone (NCBI:P:68249)
R:Edwards, M.C.; Gibbs, R.A. Genomics 14, 590-597, 1992
A:Title: A human dimerism resulting from loss of an Alu.
A:Reference number: I54176; MUID:93052387; PMID:1330888
A:Accession: I54176
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-72 <RES>
A:Cross-references: GB:U47924; GB:M6525; GB:U72506; NID:91633547; PIDN:AA51309.1; PID: A:Residues: 1-264, 'W', 266-458 <RE2>
A:Cross-references: GB:M35160; NID:9179143; PIDN:AA16069.1; PID:9179144
A:Comment: Macrophage tropic strains of HIV-1 bind to a complex of chemokine (C-C) receptor C:Genetics:
A:Gene: GDB:CD4
A:Cross-references: GDB:119767; OMIM:186940
A:Map position: 12pter-12p12
A:Introns: 16/3
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: AIDS; duplication; glycoprotein; T-cell; transmembrane protein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-458/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>
F:134-111/Domain: immunoglobulin homology <IM1>
F:136-186/Domain: immunoglobulin homology #status atypical <IM2>
F:216-299/Domain: immunoglobulin homology <IM3>
F:321-372/Domain: immunoglobulin homology <IM4>
F:397-420/Domain: transmembrane #status predicted <TM4>
F:421-458/Domain: intracellular #status predicted <INT>
F:41-109, 155-184, 328-370/Disulfide bonds: #status experimental
F:296,325/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match	42.4 %	Score 1023;	DB 1;	Length 458;
Best Local Similarity	99.5 %	Pred. No. 3.9e-55;		
Matches 199;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	MNRGVPFRLLLVLTQALPAATQGNKVVGGKGDVETCTASQKSIOPHMKNSNOIK	60	
DB	1	MNRGVPFRLLLVLTQALPAATQGNKVVGGKGDVETCTASQKSIOPHMKNSNOIK	60	
QY	61	ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFLLINLKIKEDSDTYICEVEDQKEEVQL	120	
DB	61	ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFLLINLKIKEDSDTYICEVEDQKEEVQL	120	
QY	121	LVFGLTANDTHLLOQCSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG	180	
DB	121	LVFGLTANDTHLLOQCSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG	180	

QY 181 TWTCTVLQKQKVEFKIDIV 200
 |||||
 Db 181 TWTCTVLQKQKVEFKIDIV 200

RESULT 11

G3MSM
 Ig gamma-3 chain C region, membrane-bound form - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 13-Aug-1986 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: A02156; A02155
 R:Wells, J.A.; Word, C.J.; Rimm, D.; Der-Bajan, G.P.; Martinez, H.M.; Tucker, P.W.; Blatt
 EMOB J. 3, 2041-2046, 1984
 A:Title: Structure analysis of the murine IgG3 constant region gene.
 A:Reference number: A02156; MUID:850277161; PMID:6092053
 A:Accession: A02156

A:Molecule type: DNA

A:Residues: 1-398 <WEL>

A:Cross-references: GB:J00451; NID:9194392; PIDN:AA859655.1; PID:9194433

A>Note: the sequence was determined from the germline gene

R:Komaromy, M.; Clayton, L.; Rogers, J.; Robertson, S.; Kettman, J.; Wall, R.

Nucleic Acids Res. 11, 6775-6785, 1983

A:Title: The structure of the mouse immunoglobulin in gamma-3 membrane gene segment.

A:Reference number: A02155; MUID:84041483; PMID:6314258

A:Accession: A02155

A:Molecule type: DNA

A:Residues: 328-332, 'G', 334-341, 'Q', 343-387, 'F', 389-398 <KOM>

A:Cross-references: GB:K00688

A>Note: the sequence was determined from the germline gene

C:Genetics:

A:Introns: 97/1; 113/1; 223/1; 328/1; 371/3

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob

F:19-83/Domain: immunoglobulin homology <IM1>

F:97-112/Region: hinge

F:136-205/Domain: immunoglobulin homology <IM2>

F:242-309/Domain: immunoglobulin homology <IM3>

F:346-362/Domain: transmembrane #status predicted <TM>

F:363-398/Domain: intracellular #status predicted <INT>

F:179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 39.6%; Score 955.5; DB 1; Length 398;

Best Local Similarity 61.4%; Pred. No. 4,3e-51;

Matches 180; Conservative 40; Mismatches 58; Indels 15; Gaps 4;

170 SVSOLELQSGTW-----TCTVLQKQKVEFKIDIV---VPCAPAPPKSCDKHTGCP--EL 219

62 SLSSLVTVSSSTWPSQTVCNVAHPASKTELKRLPRIPKSTPPGSG-----SCPNGI 116

220 LGSPSVFLPPEPKDTLMISRTPEVTCVVDVSHEDPEVKFMVYDGVENVNAKTKPREE 279

117 LGSPSVFLPPEPKDTLMISRTPEVTCVVDVSHEDPEVKFMVYDGVENVNAKTKPREE 176

280 QYNSTRVSVLTVDHQLNGEKYCKSNKALPAPIKTIISKAKGQPREPOVYTLPPS 339

177 QYNSTRVSVLTVDHQLNGEKYCKSNKALPAPIKTIISKAKGQPREPOVYTLPPS 236

340 RELTQNGVSLCLVGVGFPSPDIAYEWESNGQPENNYKTPVLDSDGFFLYSKLTVDX 399

237 KQWMSKKKSLTCLVTNFSSEAIIVEMKNGELSDQYKTPILSDGFFLYSKLTVDX 296

400 SRMOQNVFSCVMEHALNHYTOKSLSLSPGLQDETCAEAQDGLDGLMTT 452

297 DSNLQGEIRFTCSVMEHALNHYTOKSLSLSPGLQDETCAEAQDGLDGLMTT 349

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

C:Date: 17-Dec-1982 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999

C:Accession: B02159; A02160; B02158

R:Honojo, T.; Obara, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takahashi, N.;

Cell 18, 559-568, 1979

A:Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain g

A:Reference number: A02159; MUID:80045036; PMID:115593

A:Accession: B02159

A:Molecule type: DNA

A:Residues: 1-393 <HON>

A:Cross-references: GB:J00453

A>Note: the sequence was determined from the germline gene

R:Tyler, B.M.; Coman, A.F.; Gerondakis, S.D.; Adams, J.M.; Bernard, O.

Proc. Natl. Acad. Sci. U.S.A. 79, 2008-2012, 1982

A:Title: mRNA for surface immunoglobulin gamma chains encodes a highly conserved transme

A:Reference number: A02160; MUID:82197626; PMID:6804950

A:Accession: A02160

A:Molecule type: mRNA

A:Residues: 323-393 <TYL>

R:Rogers, J.; Choi, E.; Souza, L.; Carter, C.; Word, C.; Kuehl, M.; Eisenberg, D.; Wall,

Cell 26, 19-27, 1981

A:Title: Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma c

A:Reference number: A02158; MUID:82115295; PMID:6795207

A:Accession: B02158

A:Molecule type: DNA

A:Residues: 323-366 <ROG>

A>Note: this sequence is the translation of the first exon of the M segment

C:Comment: Cell lines producing IgG contain two mRNA species for Ig gamma chains. The ma

nucleotide membrane-bound chains in that it contains an alternative 3' end, encoded in separa

C:Genetics:

A:Introns: 1/1; 98/1; 111/1; 218/1; 323/1; 366/3

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob

F:131-200/Domain: immunoglobulin homology <IM1>

F:340-357/Domain: transmembrane #status predicted <TM>

F:358-393/Domain: intracellular #status predicted <INT>

F:174,278/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 38.6%; Score 932.5; DB 1; Length 393;

Best Local Similarity 55.0%; Pred. No. 1.1e-49;

Matches 177; Conservative 50; Mismatches 60; Indels 35; Gaps 5;

133 LLQGSLLTLTSPSSPSVQCSRPGRKNIQGGKTLVSQLELQDSTGWTCTVLQ--NQ 190

56 VLQSDLYLTSSSVTVSSP-----RPSTVTCVAHPASS 90

191 KKEVEFKIDIVPCAPAPPKSCDKHTGCPPELLGSPSVFLPPEPKDTLMISRTPEVTCVVD 250

91 TKVDKATVPRQCG-KPCIC-----TVPEV---SSVFIRPPPKDVTTLTPKVTGVVD 142

251 VSHEDPEVKFMVYDGVENVNAKTKPREQYNSTRVSVLTVDHQLNGEKYCKVSN 310

143 ISKDDPEVQFWMFVDVETHAQIQPREQYNSTRVSVLTVDHQLNGEKYCKVSN 202

311 KALPAPIKTIISKAKGQPREPOVYTLPPSRBELTKNQVSLCLVKGFPSPDIAYEWESNG 370

203 AAFPAPIKTIISKAKGQPREPOVYTLPPSRBELTKNQVSLCLVKGFPSPDIAYEWESNG 262

371 QPENNYKTPVLDSDGFFLYSKLTVDKSMQOQNVFSCVMEHALNHYTOKSLSLSP 430

263 QPAENYKTPQIPMTNGSYFYSLKLVNOKSNWEAGNTFTCSVLHGLNHTTKSLSHSP 322

431 GLQDETCAEAQDGLDGLMTT 452

323 GLQDETCAEAQDGLDGLMTT 344

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

RESULT 13

S22080
 Ig heavy chain precursor (B/MT. 4A.17.H5.A5) - bovine
 N:Alternate names: Ig gamma-1 chain C region (clone 8.10)
 C:Species: Bos primigenius taurus (cattle)

C>Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text_change 23-Jul-1999
 C/Accession: S22080; S06610; A31303
 R/Sanders, P.G.
 submitted to the EMBL Data Library, November 1991
 A/Reference number: S22080
 A/Accession: S22080
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-470 <SAS>
 A/Cross-references: EMBL:X62916; NID:q439; PIDN:CAA44639.1; PID:q440
 R/Symons, D.B.A.; Clarkson, C.A.; Beale, D.
 Mol. Immunol. 26, 841-850, 1989
 A/Title: Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma 2
 A/Reference number: S06610; MUID:90097956; PMID:2513487
 A/Accession: S06610
 A/Molecule type: DNA
 A/Residues: 142-470 <SYM>
 A/Cross-references: EMBL:X16701
 A/Note: the sequence was determined from the germline gene
 C/Genetics:
 A/Introns: 98/1; 111/1; 221/1
 C/Superfamily: Immunoglobulin C region; immunoglobulin homology
 C/Keywords: glycoprotein; heterotetramer; immunoglobulin; membrane protein
 F/161-225/Domain: immunoglobulin homology <IMM>
 F/318/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 37.9%; Score 915.5; DB 2; Length 470;
 Best Local Similarity 43.6%; Pred. No. 14e-48;
 Matches 214; Conservative 52; Mismatches 136; Indels 89; Gaps 15;

Oy	6	PERHLLVLO-----LALLPATQGNKVLGKGGDTVLTCTAS--QKKSIOFMKNSNQ	58
Db	3	PLMTLLFLVLSAPIGVLSQVQVRESGPLV--KPSQTLTLCTVSGFSLSALTMVRAP	60
Oy	59	IKILNGSGFLTKGPSKLNDRADRSRLMDGN--PLIIRKLEIDSDTYIC-----	109
Db	61	GKALEWVGITSGGTYVNPALKSRLSITKENSQSQVLSVSVPEPTATYCARSTYG	120
Oy	110	EVEDQKEVQVLLFGLTNSDTHLLQGSILTLTLESPGSSPSVQ-----CRSPRGKNIQ	164
Db	121	EVGD-----GALIDA-----WGQGLVTVSSASTAPKVPYPLSSCGGDKSSSTVT	165
Oy	165	GGKTLVSQLE-----LQDSG-----TWTCVT	186
Db	166	LGCLVSVMPPEVYVTMNSGALKSGVHTFPAVLQSGGYSLSMWTVGTSQGFICNV	225
Oy	187	LQ--NOKVVEFKIDIVPCPAPBPKSCDKTHTC--PELLGSPSVLPFPKPYDTLMISRT	242
Db	226	AHPASSTVVDKAVD---PTCKPSPCD---CCPPPELPGSPSVTFPFPKPYDTLISGTP	278
Oy	243	EVTGVVVVSHEDPEVKFNWYVDGVEVHNATKPREEOYNSYRVVSLVTLVHODMNGK	302
Db	279	EVTGVVVVGHDDPEVKFNFVDVENVATITKPEEEOFNSTYRVVSLRLRQHOOMTGGK	338
Oy	303	EYKCVSNKALPAPLEKTSKAKQPREPOVYTLTPPSDELTQKQVSLTCLVKGFPYSDI	362
Db	339	EFKCVNHEGLPAPLVRTISRTKGPAPRPQVYVLAPEQELSKSVSLTCLVTSYTPYI	398
Oy	363	AVENESNGQP--ENNYKTTTPVLDSDGSPFLYSKLTVDKSRMQQGNVFCSCVMHEALNH	420
Db	399	AVEMORNQPESEDEKGTTPQLDADSSYFLYSKLRVDRNSWQEGDYTCVVMHEALNH	458
Oy	421	YTKSLSLSPG 431	
Db	459	YTKSTSKSAG 469	

RESULT 14
 GZMSAM
 Ig gamma-2a chain C region, membrane-bound form - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 19-Feb-1984 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999

C/Accession: A02154; B32657; I57809
 R/Yamawaki-Kataoka, Y.; Nakai, S.; Miyata, T.; Honjo, T.
 Proc. Natl. Acad. Sci. U.S.A. 79, 2623-2627, 1982
 A/Title: Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin
 A/Reference number: A02154; MUID:6222190; PMID:6283537
 A/Accession: A02154
 A/Molecule type: DNA
 A/Residues: 329-399 <YAM>
 A/Cross-references: GB:J00471
 A/Note: the sequence was determined from the germline gene
 R/Yamawaki-Kataoka, Y.; Miyata, T.; Honjo, T.
 Nucleic Acids Res. 9, 1365-1381, 1981
 A/Title: The complete nucleotide sequence of mouse immunoglobulin gamma 2 A gene and env
 A/Reference number: A32657; MUID:81198976; PMID:6262729
 A/Accession: B32657
 A/Molecule type: DNA
 A/Residues: 1-329, 'K' <YAM>
 R/Hall, B.; Milcarek, C.
 Mol. Immunol. 26, 819-826, 1989
 A/Title: Sequence and polyadenylation site determination of the murine immunoglobulin gamma
 A/Reference number: I57809; MUID:90097953; PMID:2513486
 A/Accession: I57809
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 373-399 <RES>
 A/Cross-references: GB:M35032; NID:q194478; PIDN:AAA37919.1; PID:g387217
 C/Comment: The sequence of residues 1-328 was assumed to be identical with the correspond
 hat it contains an alternative 3' end, encoded in separate exons, that is homologous with
 C/Genetics:
 A/Introns: 1/1; 98/1; 114/1; 224/1; 329/1; 372/1
 C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
 C/Superfamily: immunoglobulin C region; immunoglobulin homology
 C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobul
 F/137-206/Domain: immunoglobulin homology <IMM>
 F/346-393/Domain: transmembrane #status predicted <TM>
 F/364-399/Domain: intracellular #status predicted <INT>
 F/180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 37.6%; Score 906.5; DB 1; Length 399;
 Best Local Similarity 60.5%; Pred. No. 4.2e-48;
 Matches 173; Conservative 34; Mismatches 60; Indels 19; Gaps 5;

Oy	179	SGTW-----TCTVLO--NOKVVEFKID-----IYCPAPBPKSCDKTHTCPELLGSPVF	226
Db	72	SSTWPSQITCNVNAHPASSTKVDKLEPRGPTIKPC---PKCC---PAPNLGGFSVF	124
Oy	227	LFPKPYDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNATKPREEOYNSYR	286
Db	125	IFPKIKDVLMISLPIVTCVVVDVSEDDPDVQISFVNNVENVTAQTQTHREDYNSTLR	184
Oy	287	VSVSLVTLVHODMNGKVKYKQVSNKALPAPLEKTSKAKQPREPOVYTLTPPSDELTQK	346
Db	185	VSVSLPIQHDMMWKGKFKKQVNNKDLPALEKTSKPKSVAPQVYVLPPEEEMTKK	244
Oy	347	QVSLTCLVKGFPYSDIAVENESNGQPENNYKTTTPVLDSDGSPFLYSKLTVDKSRMQQGN	406
Db	245	QVTLTCNVTPMPEPDIVENVNTNGKTELNTKTEPVLDSGYSYMYSKLVEKQNVERN	304
Oy	407	VFSCSVNHEALNNHYTKSLSPGLQLDETCAEADQDELGLMTT	452
Db	305	SYSCSVVHEGLNNHTTKSPSRFGLDLDVCAEADQGEIDGLMTT	350

RESULT 15
 I47159
 Ig gamma 2a chain constant region - pig (fragment)
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 21-Jan-2000
 C/Accession: I47159
 R/Kachkovich, I.; Sun, J.; Butler, J.E.
 J. Immunol. 153, 3565-3573, 1994

A:Title: Five putative subclones of swine Igg identified from the cDNA sequences of a

A:Reference number: 147158; MUID:95015845; PMID:7930579

A:Accession: 147159

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-328 <KAC>

A:Cross-references: EMBL:U03779; NID:9433123; PIDD:AAA52217.1; PID:9433124

C:Gene: IgG2a

C:Superfamily: Immunoglobulin C region; Immunoglobulin homology

F.133-202/Domain: Immunoglobulin homology <IMW>

Query Match 37.3%; Score 901; DB 2; Length 328;

Best Local Similarity 58.5%; Pred. No. 7.2e-48;

Matches 182; Conservative 38; Mismatches 57; Indels 34; Gaps 9;

QY 137 QSLTLESPPGSS-----PSVQCRSPRG-KNIGGKTLSSQLELDQSGTWCTVYQNG 190

Db 35 EPVATWNSGALLSSGVHTFPVSV--LQPSGLYLSMSVTPASSL--SSKSYTCNV--NH 87

QY 191 KKVEKID-----IVPCPAPEPKSCDKHTTCEPLIG-GPSVFLPPEPKPDITMISRT 242

Db 88 PATTTKVDKRGVTKTKPCP-----ICPACSPGCSVFIPEPKPDITMISRT 136

QY 243 EVTCVVDVSHDEPVEKFNMYDGVGVHNAKTKPREQYNSTRVSVYLVTHQDWLNGK 302

Db 137 QVTCVVDVSOENPEVQPSWYVDGVEVHTAOTRPREQFNSTRVSVLPIDQDWLNGK 196

QY 303 EYKCVNSKALPAPEKTIISAKGQPREPOVYTLPPRSDELTKNQVSLTICVKGFPYSDI 362

Db 197 EFKCVNNKMDPAPTRIISAKGQTRPOVYTLPPHAEELSRKSTVTCIVIFPEPDI 256

QY 363 AVEMESNGQ--PENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNVSCSMVEALHNH 420

Db 257 DVENQNRNQPREPEGYRTTPQGDVDGYTFLYSKFSVDKASMOGGITQCAVMEALHNH 316

QY 421 YTKSLSLSPG 431

Db 317 YTKSISKTPG 327

RESULT 16

GHRB

I9 gamma chain C region - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 24-Apr-1984 #sequence revision 15-Nov-1984 #text change 16-Jul-1999

C:Accession: A01749; A90290; A93928; A90245; A94416; A02161

R:Bernstein, K.E.; Alexander, C.B.; Mage, R.G.

I:Immunogenetics 18, 387-397, 1983

A:Title: Nucleotide sequence of a rabbit Igg heavy chain from the recombinant F-I haplotype

A:Reference number: A01749; MUID:84030930; PMID:6313520

A:Accession: A01749

A:Molecule type: mRNA

A:Residues: 1-323 <BER>

A:Note: This sequence has the d12 allotypic marker, 104-Thr, and the e14 marker, 185-Thr

R:Pratt, D.M.; Mole, L.E.

Biochem. J. 151, 337-349, 1975

A:Title: Sequence studies on the constant region of the Pd sections of rabbit immunoglobulin

A:Reference number: A90290; MUID:76135469; PMID:1243651

A:Accession: A90290

A:Molecule type: Protein

A:Residues: 1-47, 'E', '49-71, 'PV', '72-128 <PRA>

R:Martens, C.L.; Moore, K.W.; Steimetz, M.; Hood, L.; Knight, K.L.

Proc. Natl. Acad. Sci. U.S.A. 79, 6018-6022, 1982

A:Title: Heavy chain genes of rabbit Igg; Isolation of a cDNA encoding gamma heavy chain

A:Reference number: A93928; MUID:83299917; PMID:6193512

A:Accession: A93928

A:Molecule type: mRNA

A:Residues: 88-103, 'M', '105-143, 'E', '145-184, 'A', '186, 'E', '188-266 <MAR>

A:Cross-references: GB:M6426; NID:g165111; PIDD:AAA31289.1; PID:g165112

A:Note: This sequence has the d11 allotypic marker, 104-Met, and the e15 allotypic marker

R:Fincher, R.G.; Jackson, S.A.; Mole, L.E.; Porter, R.R.

Biochem. J. 116, 249-259, 1970

```

A>Title: Sequence studies of the Pd section of the heavy chain of rabbit immunoglobulin
A:Reference number: A90245; MUID:70110015; PMID:5461106
A:Accession: A90245
A:Molecule type: protein
A:Residues: 132-143 'E',145-161 <PRU>
R:Hall, R.L.; Lebovitz, H.E.; Fellows Jr., R.E.; Delaney, R.
In Gamma Globulins, Nobel Symp. 3, Kiliander, J., ed., pp.109-127, Almqvist and Wiksell,
A:Reference number: A94416
A:Accession: A94416
A:Molecule type: protein
A:Residues: 129-131,155-172, 'D',174-184 'A',186,'E',188-200, 'D',202-217, 'E',219-232, 'Q',
A>Note: This has the e15 allotypic marker, 185-Ala
C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoproteins; heterotrimer; immunoglobulin
F:20-82/Domains: immunoglobulin homology <IM>
F:130-199/Domains: immunoglobulin homology <IM2>
F:236-303/Domains: immunoglobulin homology <IM3>
F:173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      37.2%   Score 899; DB 1; Length 323;
Best Local Similarity 61.0%; Pred. No. 9.4e-48;
Matches 178; Conservative 34; Mismatches 64; Indels 16; Gaps 5;

QY      142 TLSPSSPSVQCRRSRGNKIQGKTLVSOLELDPDSGTCTVTVLQNOKVEFKIDIVP 201
         ||:::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB       45 TLTNGVRTPFPVRGS-----GLYSLSVSVSTSSGVTCNVAA--HPATNTKVD--- 92

QY      202 CPAPBPSCDKHTHC--PELLGPVSFLFPKKPKDTLMISRTBRVTCVVVDVSHEDPEVK 259
         ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB       93 -KTVAPEPTCSKP--TCCPELLGGPSVFIPFPKPKDTLMISRTREVTCCVVVDSDDEPEVQ 150

QY      260 FMYVVDGEVYNATKRREDOYSTRVSVLVLTVLHODMLNGRKCKCAENSKALPAPIEK 319
         ||:::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB       151 FTWITNNEQVTKARPLRLBEQGFNSTLRVSTLPITPHODWLRGAEFKCKAHNKALPAPIEK 210

QY      320 TISKAKQPPEPPQYVTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNGOPENNYKIT 379
         ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB       211 TISKARQGLEPKKYTMGPREFELSSRSVALTLCINGFIYSDISVENEGKAGKADENVKTT 270

QY      380 PPVIDSGSFFLYSKLTVDKSRWQGNSFGSCVMHEALHNHYTQKSISLSPG 431
         ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB       271 PAVIDSQSYFLYLNKLSTVPTESEWGQGDVFTFCVMEHALHNHYTQKSISRSPG 322

RESULT 17
147162
I9 gamma 4 chain constant region - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 21-Jan-2000
C:Accession: I47162
R:Kacskovics, I.; Sun, J.; Butler, J.E.
J:Immunol. 153, 3565-3573, 1994
A>Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a s
A:Reference number: I47158; MUID:95015845; PMID:7930579
A:Accession: I47162
A>Status: preliminary; translated from GB/EMBL/DDBI
A:Molecule type: mRNA
A:Residues: 1-277 <KaC>
A:Cross-references: EMBL:U03782; NID:9431129; PIDN:AAA52220.1; PID:9433130
C:Gene: IgG4
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:82-151/Domains: immunoglobulin homology <IM>

Query Match      37.2%   Score 898; DB 2; Length 277;
Best Local Similarity 62.8%; Pred. No. 9e-48;
Matches 172; Conservative 33; Mismatches 43; Indels 26; Gaps 6;

QY      168 TLSVSOLELDPDSGTCTVTVLQNOKVEFKTD-----IVCPAPBPSCDKHTTCEPLL 220
         ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB       19 TVPASSL-----SSKSYTCNV--NHPATTTKDDKVGTKRPCC-----ICPACE 62

```

```

QY 221 G-GPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREE 279
DB 63 GGGPAPAFIPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREE 122
QY 280 QYNSTYRVVSVLTVTHQDMLNKEEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPS 339
DB 123 QFNSTYRVVSVLTPIQHOMLNGEKKFCKVNNKDLPAPIRITISKAKGQTRFPOVYTLPP 182
QY 340 RDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQ--PENNYKTPPVLDSDGSFFLYSKLT 397
DB 183 TEELSRKSKYTLTCLVKGFPYPSDIAVEMESNGQPEPEGVNRTITPQDDVDGTYFLYSKLA 242
QY 398 DKSRMOQGNVFCSCVNHHEALHNHYTOKSLSPG 431
DB 243 DKASMOGQGTFCQAVMHHEALHNHYTOKSLSPG 276

RESULT 18
147160
IG gamma 2b chain constant region - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 21-Jan-2000
C:Accession: 147160
R:Kacskovics, I.; Sun, J.; Butler, J.E.
J. Immunol. 153, 3565-3573, 1994
A:Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a
A:Reference number: 147158; MUID:95015845; PMID:7930579
A:Accession: 147160
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-328 <KAC>
A:Cross-references: EMBL:U03780; NID:g433125; PIDN:AAA52218.1; PID:g433126
C:Genetics:
A:Gene: IgG2b
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:133-202/Domain: immunoglobulin homology <IMM>

Query Match 37.2%; Score 898; DB 2; Length 328;
Best Local Similarity 62.8%; Pred. No. 1.1e-47;
Matches 172; Conservative 33; Mismatches 43; Indels 26; Gaps 6;

168 TLVSQLELDSDGWTCTVYLQNKVEFKID-----IVCPAPEPKSCDKHTCPELL 220
DB 70 TVPASL---SSKSYTCNV--NHPTTKVDKRGVGTKTPCP-----ICPACE 113
QY 221 G-GPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREE 279
DB 114 SPGPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREE 173
QY 280 QYNSTYRVVSVLTVTHQDMLNKEEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPS 339
DB 174 QFNSTYRVVSVLTPIQHOMLNGEKKFCKVNNKDLPAPIRITISKAKGQTRFPOVYTLPPH 233
QY 340 RDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQ--PENNYKTPPVLDSDGSFFLYSKLT 397
DB 234 AEELSRKSKYTLTCLVKGFPYPSDIAVEMESNGQPEPEGVNRTITPQDDVDGTYFLYSKPSV 293
QY 398 DKSRMOQGNVFCSCVNHHEALHNHYTOKSLSPG 431
DB 294 DKASMOGQGTFCQAVMHHEALHNHYTOKSLSPG 327

RESULT 19
GGGP
IG gamma-2 chain C region - guinea pig
C:Species: Cavia caprellus (guinea pig)
C:Date: 07-May-1981 #sequence_revision 07-May-1981 #text_change 16-Jul-1999
C:Accession: A94553; A90352; A90359; A90384; A90385; A02151
R:Trischmann, T.M.
submitted to the Atlas, April 1975
A:Reference number: A94553
A:Accession: A94553

```

```

A:Molecule type: protein
A:Residues: 1-3 <TRI>
R:Bitenslein, B.K.; Hussain, O.Z.; Cebra, J.J.
Biochemistry 10, 18-25, 1971
A:Title: Structure of heavy chain from strain 13 guinea pig immunoglobulin-G(2). III. Am
A:Reference number: A90352; MUID:71058471; PMID:5538606
A:Accession: A90352
A:Molecule type: protein
A:Residues: 4-68 <BIR>
R:Turner, K.J.; Cebra, J.J.
Biochemistry 10, 9-17, 1971
A:Title: Structure of heavy chain from strain 13 guinea pig immunoglobulin-G(2). II. Amir
A:Reference number: A90359; MUID:71058486; PMID:5538616
A:Accession: A90359
A:Molecule type: protein
A:Residues: 69-133;312-329 <TUR>
R:Tracey, D.E.; Cebra, J.J.
Biochemistry 13, 4796-4803, 1974
A:Title: Primary structure of the C-H2 homology region from guinea pig IgG2 antibodies.
A:Reference number: A90384; MUID:75036072; PMID:4429665
A:Accession: A90384
A:Molecule type: protein
A:Residues: 134-226 <TRA>
R:Trischmann, T.M.; Cebra, J.J.
Biochemistry 13, 4804-4811, 1974
A:Title: Primary structure of the C-H3 homology region from guinea pig IgG2 antibodies.
A:Reference number: A90385; MUID:75036073; PMID:4609467
A:Accession: A90385
A:Molecule type: protein
A:Residues: 227-311 <TR2>
R:Oliveira, B.; Lam, M.E.
Biochemistry 10, 26-31, 1971
A:Title: Interchain disulfide bridges of guinea pig gamma-2- immunoglobulin.
A:Reference number: A90354; MUID:71058474; PMID:4922544
A:Contents: annotation; disulfide bonds
A>Note: Cys-16 is involved in a heavy-light chain bond
A>Note: Cys-105, Cys-107, and Cys-110 form inter-heavy chain bonds
C:Comment: This chain was isolated from pooled serum of strain 13 inbred guinea pigs.
C:Complex: An immunoglobulin heterodimer subunit consists of two identical light (kap)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterodimer; immunoglobulin
F:21-81/Domain: immunoglobulin homology <IM1>
F:135-204/Domain: immunoglobulin homology <IM2>
F:241-310/Domain: immunoglobulin homology <IM3>
F:28-79/Disulfide bonds: #status experimental
F:142-202/Disulfide bonds: #status experimental
F:178/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:248-308/Disulfide bonds: #status experimental

Query Match 37.1%; Score 894.5; DB 1; Length 329;
Best Local Similarity 62.5%; Pred. No. 1.8e-47;
Matches 172; Conservative 34; Mismatches 60; Indels 9; Gaps 4;

163 IGGKTLVSQLELDSDGWTCTVYLQ--NKKVEFKIDIVCPAPEPKSCDKHTC--PE 218
DB 58 LQSLGLVSLTSMVTPSSQKATCNVAHPASSTKYDTEVPIRTPBECTCPK---CPPE 114
QY 219 LIGGPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREE 278
DB 115 NLGGPSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREE 174
QY 279 EQNSTYRVVSVLTVTHQDMLNKEEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPP 338
DB 175 EQYNTFRVSVLTPIQHOMLNGEKKFCKVNNKDLPAPIEKTISKAKGAPMPVYTLPP 234
QY 339 SRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQ--ENNYKTPPVLDSDGSFFLYSKLT 396
DB 235 SRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPEPEGVNRTITPQDDVDGTYFLYSKLT 294
QY 397 VDKSRMOQGNVFCSCVNHHEALHNHYTOKSLSPG 431
DB 295 VDKSAMDOGTYTCSVNHHEALHNHYTOKSLSPG 329

```

RESULT 20

Ig gamma 1 chain constant region - pig (fragment)
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 21-Jan-2000
 C:Accession: I47158
 R:Kakekoyev, I.; Sun, J.; Butler, J.E.
 J. Immunol. 153, 3565-3573, 1994
 A:Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a B cell cDNA library
 A:Reference number: I47158; MUID:95015845; PMID:7930579
 A:Accession: I47158
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-328 <YAC>
 A:Cross-references: EMBL:U03778; NID:9433121; PIDN:AAA52216.1; PID:9433122
 C:Genetics:
 A:Gene: IgG1
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 F:133-202/Domain: immunoglobulin homology <IMM>

Query Match 36.4%; Score 879.5; DB 2; Length 328;
 Best Local Similarity 55.2%; Pred. No. 1.5e-46;

Matches 180; Conservative 38; Mismatches 67; Indels 41; Gaps 7;

```

QY 147 PSSSPVQCRSPRGKNIQGGKT---LSVQLELDSDGTWC-----TYLQNK 191
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 2 PRTAPSVYPLAPGSDVSPNVALGCLASSYPEPEVTVNMGALTVGHTPEPSVLPQSG 61
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 192 KVEFKLDI-VPCAPPEKKS--CDKTH-----TCPELG-----GPSVL 227
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 62 LYSLSMTVTPVASSLSKSKYTCTVNVNRPATTKVXKVGIGHORPTCDICGCEVAGPSVTI 121
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 228 PPPPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDSEVHNATKPREBOYNSTYRV 287
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 122 PPPPKDTLMISQTPETVTCVVVDVSKHALEVGSMVVDGEVHTAETRKEQFNSTYRV 181
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 288 VSVLTVLHDDMLNGEKYKCKVSNKALPAPIETKISAKAQPREPOVYITLPPSRDELTKQV 347
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 182 VSVLPIQHDWMSGKFKCKKVNVDLPADITRTISVAIQSREPOVYITLPPPAEELSRK 241
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 348 VSVLTVLNGEYPSDIAVENESNGQ--PENNYKTTPTVLDSDSFFLYSLTVDKSRMOQG 405
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 242 VLTCLVIGFPPDPDIHVEKMSNGQPEPENTRTTPQDDVDGTFLLYSKLAVDKAKMDHG 301
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 406 NVFSGSVMEALHNHYTKSLSPG 431
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 302 DKPECAVMEALHNHYTKSLSPG 337
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 21

IG gamma-2b chain C region, membrane-bound form - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 17-Dec-1982 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: C02154; A02158; B02157
 R:Yamawaki-Kataoka, Y.; Nakai, S.; Miyata, T.; Honjo, T.
 Proc. Natl. Acad. Sci. U.S.A. 79, 2623-2627, 1982
 A:Title: Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin A
 A:Reference number: A02154; MUID:82222190; PMID:6283537
 A:Accession: C02154
 A:Molecule type: DNA
 A:Residues: 335-405 <YMW>
 A:Cross-references: GB:J00462
 R:Rogers, J.; Choi, E.; Souza, L.; Carter, C.; Word, C.; Kuehl, M.; Eisenberg, D.; Wall, Cell 26, 19-27, 1981
 A:Title: Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma C
 A:Reference number: A02158; MUID:82115295; PMID:6799207
 A:Accession: A02158
 A:Molecule type: DNA
 A:Residues: 335-378 <ROG>
 A:Note: the translation of the first exon of the membrane-bound segment is given

R:Yamawaki-Kataoka, Y.; Kataoka, T.; Takahashi, N.; Obata, M.; Honjo, T.

Nature 283, 786-789, 1980

A:Title: Complete nucleotide sequence of immunoglobulin gamma2b chain gene cloned from a mouse B cell

A:Reference number: A02157; MUID:80120716; PMID:676534

A:Contents: a allele

A:Accession: B02157

A:Molecule type: DNA

A:Residues: 1-335, 'K' <YA2>

C:Comment: The sequence of residues 1-334 was assumed to be identical with the corresponding sequence of the gamma 2b chain gene

C:Comment: Cell lines producing IgG contain two mRNA species for Ig gamma chains. The major species is the one that contains an alternative 3' end, encoded in separate exons, that is homologous with the gamma 2b chain

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin; intracellular; status predicted <IMM>

F:143-212/Domain: immunoglobulin homology <IMM>

F:352-369/Domain: transmembrane #status predicted <IMM>

F:370-405/Domain: intracellular #status predicted <INT>

F:186/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 36.4%; Score 877.5; DB 1; Length 405;
 Best Local Similarity 51.7%; Pred. No. 2.5e-46;

Matches 178; Conservative 52; Mismatches 75; Indels 39; Gaps 9;

```

QY 133 LIQG---OSLTITLSPGSS-----PSVQCRSPRGKNIQGGKTVSVQLELDSDGTW-- 182
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 28 LVKGFPPSVTVTVNMGSLSSSVHTFPAL-----LQSG-LYTWSSSVTPSPSTWPS 77
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 183 ---TCTVQ--NQKVEFKID-----IVCPAPPEKSCDKTHC--PELLGSPVFLF 228
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 78 QTVTVSVHAPASSTTVDDKLEPSGPISTINPDP-----PCECHCKPAPNLEGGSPVTF 132
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 229 PPPPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDSEVHNATKPREBOYNSTYRV 288
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 133 PPNIQDVLMTSLTPPVTCVVVDVSEDDPDVQISWPNVNVHTAQTHREDYNSTINRV 192
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 289 VSVLTVLHDDMLNGEKYKCKVSNKALPAPIETKISAKAQPREPOVYITLPPSRDELTKQV 348
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 193 STLPIQHDWMSGKFKCKKVNVDLPADITRTISVAIQSREPOVYITLPPPAEELSRK 252
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 349 SLTCLVIGFPPDPDIHVEKMSNGQPEPENTRTTPQDDVDGTFLLYSKLAVDKAKMDHG 408
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 253 SLTCLVIGFPPDPDIHVEKMSNGQPEPENTRTTPQDDVDGTFLLYSKLAVDKAKMDHG 312
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 409 SCVMEALHNHYTKSLSPGLQDDETCAEADGELDGLMTT 452
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 313 SCVMEALHNHYTKSLSPGLQDDETCAEADGELDGLMTT 356
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 22

IG gamma-2b chain C region, membrane-bound form - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 17-Dec-1982 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: C02154; A02158; B02157
 R:Yamawaki-Kataoka, Y.; Nakai, S.; Miyata, T.; Honjo, T.
 Proc. Natl. Acad. Sci. U.S.A. 79, 2623-2627, 1982
 A:Title: Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin A
 A:Reference number: A02154; MUID:82222190; PMID:6283537
 A:Accession: C02154
 A:Molecule type: DNA
 A:Residues: 335-405 <YMW>
 A:Cross-references: GB:J00462
 R:Rogers, J.; Choi, E.; Souza, L.; Carter, C.; Word, C.; Kuehl, M.; Eisenberg, D.; Wall, Cell 26, 19-27, 1981
 A:Title: Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma C
 A:Reference number: A02158; MUID:82115295; PMID:6799207
 A:Accession: A02158
 A:Molecule type: DNA
 A:Residues: 335-378 <ROG>
 A:Note: the translation of the first exon of the membrane-bound segment is given

A:Residues: 3-399 <FOM>
 A:Note: sequence extracted from NCBI backbone (NCBIP:118332)
 C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells
 C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
 C:Keywords: duplication; glycoprotein; T-cell; transmembrane protein
 F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>
 F:1-371/Domain: extracellular #status predicted <EXT>
 F:9-86/Domain: immunoglobulin homology <IM1>
 F:111-161/Domain: immunoglobulin homology #status atypical <IM2>
 F:191-274/Domain: immunoglobulin homology <IM3>
 F:296-347/Domain: immunoglobulin homology <IM4>
 F:372-399/Domain: transmembrane #status predicted <TM>
 F:396-433/Domain: intracellular #status predicted <INT>
 F:16-84,130-159,303-345/Disulfide Bonds: #status predicted
 F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 36.2%; Score 875; DB 1; Length 432;
 Best Local Similarity 97.7%; Pred. No. 3.8e-46;
 Matches 170; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 KVLGKGGDTVELTCTASOKSIOFHMKNNOIKILGNGSFLTKGPKLNDRADRSRL 86
 DB 2 KVLGKGGDTVELTCTASOKSIOFHMKNNOIKILGNGSFLTKGPKLNDRADRSRL 61
 QY 87 WDGNPFLIKNLKIEDSDTYICEVEDQKEEVQLLVFGILTANSDTHLLQGSLTLTLESP 146
 DB 62 WDGNPFLIKNLKIEDSDTYICEVEDQKEEVQLLVFGILTANSDTHLLQGSLTLTLESP 121
 QY 147 PGSSSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLOKQKVEFKIDIV 200
 DB 122 PGSSSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLOKQKVEFKIDIV 175

RESULT 23

147161
 I9 gamma 3 chain constant region - pig (fragment)
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 21-Jan-2000
 C:Accession: I47161
 R:Kacskovics, I.; Sun, J.; Butler, J.E.
 A>Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a
 A:Reference number: 147158; MUID:95015845; PMID:7930579
 A:Accession: I47161
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-328 <KAC>
 A:Cross-References: EMBL:U03781; NID:9433127; PIDN:AAAS2219.1; PID:9433128
 C:Genetic: A:Gene: IgG3
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 F:133-202/Domain: immunoglobulin homology <IM>

Query Match 36.2%; Score 873; DB 2; Length 328;
 Best Local Similarity 57.6%; Pred. No. 3.7e-46;
 Matches 179; Conservative 37; Mismatches 61; Indels 34; Gaps 9;

QY 137 GSLTLT-----LESPGSSPVQCRSPRG-KNIQGGKTLVSQLELDGSGTWTCTVLOQ 190
 DB 35 EGVYITNNSGALTSQVHTFPPV--LQPSGLYSLSMTVTPASSL---SSKSTTCNV--NH 87
 QY 191 KKEVEKID-----IVPCPAPEPKSCDKTHTCPEL-LGSPSVFLPEPKKDTLMISRTP 242
 DB 88 PATTKTKVDKRVGTVTKTKPCP-----ICPCEVAAGPEVVFPPKPKDTLMISGTP 136
 QY 243 EYTCVVVDVSHEDPEVKRWYVDGVEVNAKTKRPEEQNTSYRVASVLTVLHODWLNKG 302
 DB 137 EYTCVVVDVSHEDPEVKRWYVDGVEVNAKTKRPEEQNTSYRVASVLTVLHODWLNKG 196
 QY 303 EYKCKVSKALPAPIEKTSKAKGQPREPOVYTLPPSDDELTKNOVSLTCLVKGYPGSDI 362
 DB 197 EFKCKVNVNDLPAPITRTITISKALIGSREPOVITLPPPAEELSRSKVTVTCLVIGYPPDI 256

QY 363 AVEMESNGQ--PENNYKTPPVLDSGDSFPLYSKLTVDKSRMOGNVFSQVMEALNHN 420
 DB 257 HVEKSNKGQPEPEENNYTTPPQDVGDGTFPLYSKLTAVDKARMDGTFECVMEALNHN 316
 QY 421 YTKSKLSLSPG 431
 DB 317 YTKSKISKTOG 327

RESULT 24

G3MSC
 I9 gamma-3 chain C region, secreted form - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 17-Mar-1987 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: B02156
 R:Wels, J.A.; Word, C.J.; Ramm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blatier
 EMBL J. 3, 2041-2046, 1984
 A>Title: Structure analysis of the murine IgG3 constant region gene.
 A:Reference number: A02156; MUID:85027161; PMID:6092053
 A:Accession: B02156
 A:Molecule type: DNA
 A:Residues: 1-329 <WEL>
 A:Cross-References: GB:J00451
 A:Note: the sequence was determined from the germine gene
 C:Genetic: A:Introns: 97/1; 113/1; 223/1
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)
 chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1st
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobul
 F:19-83/Domain: immunoglobulin homology <IM1>
 F:97-112/Region: hinge
 F:136-205/Domain: immunoglobulin homology <IM2>
 F:242-309/Domain: immunoglobulin homology <IM3>
 F:179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 35.5%; Score 857.5; DB 1; Length 329;
 Best Local Similarity 59.6%; Pred. No. 3.2e-45;
 Matches 162; Conservative 38; Mismatches 57; Indels 15; Gaps 4;

QY 170 SVSGLLELDGSGT-----TCTVLOKQKVEFKIDI--VCPAPPEPKSCDKTHTCPE--EL 219
 DB 62 SLSSLVTPSPSTWPSQGVICNVAPKATLEIKRIEPIKPSPPSS-----SCPENI 116
 QY 220 LGGSVFLPEPKKPDYTLMSRTPEVTCVVVDVSHEDPEVFNNVVDGVEVNAKTKPRE 279
 DB 117 LGGSVFLPEPKKPDYTLMSRTPEVTCVVVDVSHEDPEVFNNVVDGVEVNAKTKPRE 176
 QY 280 QYNSTYRVSVLTVLHODWLNKGKCYKCVSNKALPAPIEKTSKAKGQPREPOVYTLPPS 339
 DB 177 QYNSTYRVSVLTVLHODWLNKGKCYKCVSNKALPAPIEKTSKAKGQPREPOVYTLPPS 236
 QY 340 RDELTKNOVSLTCLVKGYPGSDIAVEMESNGQPEENNYKTPPVLDSGDSFPLYSKLTVDK 399
 DB 237 REQNSKKKVSILTCLVTNPFSEALISVEWENGELEQDYKNTPIPLDSGYFLYSKLTVD 296
 QY 400 SRMQGNVFSQVMEALNHNHYTKSLSPG 431
 DB 297 DSWLQSGIFTCVYVHEALNHNHYTKSLSPG 328

RESULT 25

PC4436
 monoclonal antibody 13-1 heavy chain - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jan-2000
 C:Accession: PC4436
 R:Akashi, S.; Kato, K.; Torizawa, T.; Dohmae, N.; Yamaguchi, H.; Kamachi, M.; Harada, A.
 Biochem. Biophys. Res. Commun. 240, 566-572, 1997
 A>Title: Structural characterization of mouse monoclonal antibody 13-1 against a porphyr
 A:Reference number: JC5810; MUID:98063277; PMID:9358605
 A:Accession: PC4436
 A:Molecule type: protein

A:Residues: 1-444 <AKA>

C:Comment: This catalytic antibody has peroxidase oxidase activity. It is directed again

C:Superfamily: immunoglobulin C region; immunoglobulin homology

F:251-320/Domain: immunoglobulin homology <IMM>

F:22/Disulfide bonds: Interchain (to 98) #status predicted

F:99/Disulfide bonds: Interchain (to 109) #status predicted

Query Match 35.4%; Score 854.5; DB 2; Length 444;

Best Local Similarity 41.5%; Pred. No. 6.9e-45;

Matches 188; Conservative 76; Mismatches 118; Indels 71; Gaps 14;

QY 30 LGKGDVLTCTAS--QKSIQFMW-----KSNQIKIL-----GNQSGFLTKGPS 74

DB 11 LVRPGNSLTSLCTSFSTSNRMHMLRQPGKRLMIAVITKSDNYGAKTAESRGAF 70

QY 75 KANDRADSRSLMDQGNFPLIKNLKIEDSDTYIC-----EVED 113

DB 71 TTS-RDSSKSYVLQNN-----RLREEDATATYCCRTPMVYAMDQSGTSIVSSAKT 123

QY 114 QKEVQLVFGLTANSDDLQSG-----QSLTLESPPGSSPSVQCRSPRGKNIG 165

DB 124 TPSPSVPLPAGSALQNTSNVTLGCLVKGYPPEPVITWNS--GSLSSGVHTPPA--VLD 178

QY 166 GKLTVSVOLELQDSGT-----TCTVLO--NOKVVEFKIDIVPCPAPBPSCDKHTTCE 218

DB 179 SPLYLSSSVTVPSSTWPEETVCNVAHPASSTKVDKIVPRDCG--KPCIC---TYPE 233

QY 219 LUGGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKFNWYDGEVHNATKPRE 278

DB 234 V---SSVFIFPPKPKDVLITLTPKTCVVDVDSKDPVQFSWFDVDEVTAGQPRE 290

QY 279 EGVNSTYRVVSVTLVHODMLNGEKCKVSNKALPAPLEKITSKAGQPREPVYTTP 338

DB 291 EQPNSTFRVSELPFHQMLNGEKCKVSNKALPAPLEKITSKAGQPKAPQVYTTTP 350

QY 339 SDELTKNQVSLTCLVKGFPSPDIADVESNGQPNNTTPPVLDSDSFFLYSKLTYD 398

DB 351 PRQWAKDKVSLTCLMTDFPEDITVEMQNGQPAKNYKNTQPIMDTDSYFVYSKLVNQ 410

QY 399 KSRWQGNVFCGSVMHEALHNHYTQKSLSPG 431

DB 411 KSNWEAGNTFTCSVLHEGLHNHHTKSLSPG 443

RESULT 26

S31459

Ig gamma-1 chain - sheep (fragment)

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999

C:Accession: S31459

R:Parent, S.; Nau, P.

Submitted to the EMBL Data Library, December 1992

A:Reference number: S31459

A:Accession: S31459

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-472 <PAT>

A:Cross-References: EMBL:X69797

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: immunoglobulin

F:277-346/Domain: immunoglobulin homology <IMM>

Query Match 35.4%; Score 854.5; DB 2; Length 472;

Best Local Similarity 42.2%; Pred. No. 7.5e-45;

Matches 196; Conservative 55; Mismatches 119; Indels 94; Gaps 16;

QY 36 TVELTCTAS--QKSIQFMWKN--SNQIKIGNQ--SFLTKGP---SKLMDRADSR 85

DB 34 TISVTCTIGSFLNNYGVWQAQKALEWLGSGSIDYDIDYNPVLSRLITTDTSKS 93

QY 86 LMDQGNFPLIKNLKIEDSDTYICEVEDQKE-----VOLVFGLTANSDDL 133

DB 94 -----QVSLTSTVTETDTRAVYVCARVDYDSSHAFAVASYDFMGQLISVLSASF----- 144

QY 134 LQGSULTLTPSPSSPSVQCRPKNI-----QSGKTLVSQ 173

DB 145 -----TPPKVYPLTSCGDTSSIVTLGCLVSSWYDEPVITWNSALSGVHT 193

QY 174 LE--LQDSG-----TWCTVLO--NOKVVEFKIDIVPCPAPBPSCD 211

DB 194 PFAIIQSGSLVSSVTVPASTGAGTFFICNVAPASTKVDKVE--PGCDP--KPCIC-- 250

QY 212 KTHTC--PELLGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKFNWYDGEV 269

DB 251 ---RCPPPELPQGPVFLFPKPKDVLITLTPKTCVVDVSHEDPEVKFNWYDGEV 307

QY 270 HNAKTPREGVNSTYRVVSVTLVHODMLNGEKCKVSNKALPAPLEKITSKAGQPR 329

DB 308 RTAKTPREGVNSTYRVVSVTLVHODMLNGEKCKVSNKALPAPLEKITSKAGQPR 367

QY 330 EPQVYTLTPSRDELTKNQVSLTCLVKGFPSPDIADVESNGQ--ENNYKTTTPVLDSDG 387

DB 368 EQVYVLAHPQBELSKTLSTVCLVTFGYPDVIAEMQNGQPRESDKXGTTSQLDADG 427

QY 388 SFFLYSKLTVDSKRWQGNVFCGSVMHEALHNHYTQKSLSPG 431

DB 428 SYFLYSLRLVDSKRWQGNVFCGSVMHEALHNHYTQKSLSPG 471

RESULT 27

S37483

Ig gamma-2a chain - mouse

C:Species: Mus musculus (house mouse)

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Jul-1999

C:Accession: S37483

R:Ducancel, F.F.D.

Submitted to the EMBL Data Library, February 1993

A:Reference number: S37483

A:Accession: S37483

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-469 <DUC>

A:Cross-References: EMBL:X70423; NID:g406252; PIDN:CAA49868.1; PID:g406253

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: immunoglobulin

F:276-345/Domain: immunoglobulin homology <IMM>

Query Match 35.2%; Score 850.5; DB 2; Length 469;

Best Local Similarity 42.5%; Pred. No. 1.3e-44;

Matches 197; Conservative 56; Mismatches 125; Indels 85; Gaps 15;

QY 30 LGKGDVLTCTASQKSIQFMWKNQIKILNQ-----SFLTKGPSKLMRADSR 83

DB 30 LVKPGASVYKISCKASGYFTDY--IMWVQKQKQGLKMGITYPASGNTKNNENFKGK 86

QY 84 RSL--WDQGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDDLQ--GQS 138

DB 87 ATLVTDTSSSTAYMGLSLTSEDTAVYFC-----ARAGATAT--LDYWGQG 132

QY 139 LTLTLESPGSSPSVQ-----CRSPRGKNIGG-----KTL----- 169

DB 133 TLLTVSSAKTATPASYVPAPVCGDTTGSSTVGLVKGYPPEPVITLWNSSGLSGVHTP 192

QY 170 -SVSGLQLQ-----DSGTW-----TCTVLO--NOKVVEFKID--TVPCAPBP 208

DB 193 PAVLQSDLYTSSSVTVTSTWPSQSITCNVAHPASSTKVDKLEPRGPTTKPCP--PC 249

QY 209 SCDKHTCPPELLGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKFNWYDGEV 268

DB 250 KC---PAPNLLGGSVTFPPPKIKDVLMIISPLIVTCVVDVSEDDPDVQISFVNVE 305

QY 269 VHNATKPREGVNSTYRVVSVTLVHODMLNGEKCKVSNKALPAPLEKITSKAGQPR 328

DB 306 VHTAQQTQTHREDYNTSLVVALPIQHODMWSGKEFKVNNKDLPAPIERTISKPKGSV 365

QY 329 REPQVYTLPPSRDELTKNQVSLTCLVKGFPSPDIADVESNGQPNNTTPPVLDSDG 388

Db 366 RAPQVYLPPEEETKQVLTLCVTFDPEMEDIVETWNNKTELTANKTEPVLDSG 425
Qy 369 FELYSKLTVDKSRMOQGVFSCVWHEALHNHYTKSLSPG 431
Db 426 YFMYSKLVEKKWVERNYSVCSVHEGLHNHHTTKSFSRTPG 468

RESULT 28

C30554
Ig heavy chain C region - sheep (fragment)
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C/Date: 03-Mar-1989 #sequence_revision 03-Mar-1989 #text_change 21-Jan-2000
C/Accession: C30554
R/Foley, R.C.; Behn, K.J.
J. Immunol. 142, 708-711, 1989
A/Title: Isolation and sequence of sheep Ig H and L chain cDNA.
A/Reference number: A30554; MUID:89093962; PMID:2492052
A/Accession: C30554
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-308 <FOL>
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: heterotetramer; immunoglobulin
F:113-182/Domain: immunoglobulin homology <IMM>

Query Match 34.7%; Score 838; DB 2; Length 308;
Best Local Similarity 60.7%; Pred. No. 4,46-44;
Matches 164; Conservative 37; Mismatches 55; Indels 14; Gaps 7;

Qy 170 SVSQLELDG--TWCTVLQ--NOKYVEFKIDVPCAPBPKSCDKHTHC--PELLGSP 223
Db 44 SVTVTPASTSGAQTFICNVNAPASSTKYDKRVE--GCPDP--CKHC---RCPEPLPGP 97
Qy 224 SVFLPPPKPDTLIMISRTPEVTCVVDVSHDEPKFKNWYVDGVEVNAKTKPREQYNS 283
Db 98 SVFLPPPKPDTLITSGPEVTCVVDVGDDEPKFQFVNVNVEVTAATKPREQFNS 157
Qy 284 TYRVSVLTVLHODMLNGEKYCKVSKNALPAPIEKTISKAGQPREQVYTLPPSRDEL 343
Db 158 TFRVVSALPIQHODMTGKKEFKCKHNFALPAPIRTISRTKQARERQVYVLAPODEL 217
Qy 344 TKQVSLTCLVKGFPSDIAVWESNGP--ENNYKTPVLDSGSEFLYSKLTVDKSR 401
Db 218 SKSTLSVTCLVTFGFPDYIAVEMQNGQPESEDKTGITTSQLDADGSEFLYSRLKVDKNS 277
Qy 402 MQGNVFCSCVWHEALHNHYTKSLSPG 431
Db 278 MQESDTVACVWHEALHNHYTKSLSPG 307

RESULT 29

PS0018
Ig gamma-2b chain C region - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 16-Jul-1999
C/Accession: PS0018; B25941
R/Brueggemann, M.
Gene 74, 473-482, 1988
A/Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.
A/Reference number: PS0017; MUID:89232738; PMID:3149946
A/Accession: PS0018
A/Molecule type: DNA
A/Residues: 1-333 <BRU>
R/Brueggemann, M.; Freu, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.
Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986
A/Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody
A/Reference number: A25941; MUID:86287397; PMID:3016742
A/Accession: B25941
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 227-333 <BR2>
C/Genetics:

A/Introns: 96/1; 117/1; 227/1
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F:20-82/Domain: immunoglobulin homology <IMM>

Query Match 34.5%; Score 833; DB 2; Length 333;
Best Local Similarity 58.6%; Pred. No. 1e-43;
Matches 157; Conservative 38; Mismatches 53; Indels 20; Gaps 5;

Qy 179 SGTW-----TCVTLQ--NOKYVEFKID-----VPCAPBPKSCDKHTHC--PELLGSP 223
Db 70 SSTWPGVTCVNVNAPASSTKYDKERRNGIGHKCP-----TCPTCHKCPPELLGSP 124
Qy 224 SVFLPPPKPDTLIMISRTPEVTCVVDVSHDEPKFKNWYVDGVEVNAKTKPREQYNS 283
Db 125 SVFLPPPKPDTLITSGNAVTCVVDVSEEPDVQFSWVNVNVEVTAQTQPREQYNS 184
Qy 284 TYRVSVLTVLHODMLNGEKYCKVSKNALPAPIEKTISKAGQPREQVYTLPPSRDEL 343
Db 185 TFRVVSALPIQHODMNGEKYCKVSKNALPAPIEKTISKAGQVLRKQVYVLMGPTEOL 244
Qy 344 TKQVSLTCLVKGFPSDIAVWESNGPENNKTTPVLDSGSEFLYSKLTVDKSRMQ 403
Db 245 TEGVSLTCLVGLFPLPDIGVETSNCHIKNTIEPVNDSGSEFLYSKLVNERSMD 304
Qy 404 QGNVFCSCVWHEALHNHYTKSLSPG 431
Db 305 SRAFFVCSVWHEGLHNHVEKTSISRPG 332

RESULT 30

S40295
Ig gamma-2a chain (mAb735) - mouse
C/Species: Mus musculus (house mouse)
C/Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 16-Jul-1999
C/Accession: S40295
R/Klebert, S.; Kratzin, H.D.; Zimmermann, B.; Vaesen, M.; Frosch, M.; Weisgerber, C.; Bl
submitted to the EMBL Data Library, January 1993
A/Description: Primary structure of the murine monoclonal IgG2a antibody mAb735 against
A/Reference number: S40295
A/Accession: S40295
A/Molecule type: Protein
A/Residues: 1-446 <KLE>
C/Genetics:
A/Map position: 12
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: disulfide bond; glycoprotein; immunoglobulin; pyroglutamic acid
F:1-446/Product: Ig gamma-2a chain #status experimental <MAT>
F:1-117/Domain: V-D-D region <VDJ>
F:118-446/Domain: C region <CHR>
F:118-214/Domain: C1 region <CH1>
F:215-230/Region: hinge
F:231-340/Domain: C2 region <CH2>
F:341-446/Domain: C3 region <CH3>
F:360-427/Domain: immunoglobulin homology <IMM>
F:1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental
F:22-96,144-199,261-321,367-425/Disulfide bonds: #status predicted
F:113/Disulfide bonds: interchain (to light chain) #status predicted
F:224,227,229/Disulfide bonds: interchain #status predicted
F:297/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 34.2%; Score 825; DB 2; Length 446;
Best Local Similarity 40.9%; Pred. No. 4,4e-43;
Matches 190; Conservative 60; Mismatches 123; Indels 92; Gaps 15;

Qy 30 LGKAGDVELTCTASQKSLQF--HMKNSQIKLNGQ-----SFLTGPSKLANRAD 81
Db 11 LVPRGASVTKISCKASGYTFDYIHW-----VQRPDEGEWIGMIVPGSGNTKYNEFK 65
Qy 82 SRSL---MQGNPPLIIKXLTEDSDTYICEVDQKEVQLVFGVTAUSDTHLQ--G 136
Db 66 GKATLTVDTSSTAYWQLSSLTSEDSAVYFC-----ARGKAFMDYWG 108

```

Qy 137 QSLTLTLESPGSSPSVQ-----CRSPRGKNIQGG-----XTL----- 169
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 QGTSVTVSSAKTAPAVVPLAPVCSPTTSSVTLGLVVGFPPEPTLTMNSSLSSGVH 168
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 170 ---SVSQLEIQ-----DSGTW-----TCTVLQ--NOKRVEFKID-----IVPCPAPE 206
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 169 TEPAVLQSDLYTLSSSVTVTSSSTWPSQSIITCNVAHNASSTKVDKIEPRGPIKPCP--- 225
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 207 PKSCDHTHCPELLGGPSVFLPPPKDPLMTSRTEPVYCVVVDVSHEDPEVKFENNYVDG 266
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 226 PCKC-----PAPNLLGGPSVFIPEPKIKDVLMTLSLSPWTCVVDVSEDDPDVQISFVXN 281
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 267 VEYHNAKTRPREQVSTYRVVSVLTVLHODMLNGEYCKVSNKALPAPIEKTISKAG 326
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 282 VEVLTAGTQTHREDVNSTLRVVSALPIQHDMMKSGEFCKKNNKDLPAPIEKTISKPG 341
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 327 QPREPOVYTLPPSRDELTKNQVSLTCLVKGFPSPDIAVWESNGQPENNYKTTTPVLDSD 386
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 342 SYRAPQVYVLPPEEEMTKQVTLTCMVTDFMPEDIYVMWTNNGKTELNYKTEPVLDS 401
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 387 GSEFLYSKLTVDKSRWQGNVPSGSVMHEALHNHYOKSLSLSPG 431
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 402 GSYFMTSKLRVEKKNWERNYSYCSVYHGLHNHTTKSFSPRTPG 446
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 31

```

PS0017
Ig gamma-1 chain C region - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 16-Jul-1999
C/Accession: PS0017; C25941
R:Brueggemann, M.
Gene 74, 473-482, 1988
A/Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.
A/Reference number: PS0017; MUID:89232738; PMID:3149946
A/Accession: PS0017
A/Molecule type: DNA
A/Residues: 1-326 <BRU>
R:Brueggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.
Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986
A/Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody
A/Reference number: A25941; MUID:86287397; PMID:3016742
A/Accession: C25941
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 220-326 <BR2>
C/Genetics:
A/Introns: 98/1, 113/1, 220/1
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F/20-84/Domain: immunoglobulin homology <IMM>

```

```

Query Match 34.1%; Score 823.5; DB 2; Length 326;
Best Local Similarity 50.0%; Pred. No. 3.8e-43;
Matches 163; Conservative 53; Mismatches 55; Indels 55; Gaps 9;

Qy 133 LLQG--GSLTLTLESPGSS-----PSYQCSPPKRNIIQGGKTSLSVQLEIQDSGTW-- 182
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 28 LVKGFPEPEVTVTWSGALSSGVHFRPAV-----IQSGLYTLTSSVTV-PSSTWPS 77
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 183 ---TCTVLQ--NOKRVEFKIDIVPCAPAPKSCDHTHCPELLGG-----PSV 225
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 QTVTCVNAHPASTKVDKLI-----VPRNC-----GGDCKRCITGSEVSSV 119
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 226 FLPPPKDPLMTSRTEPVYCVVVDVSHEDPEVKFENNYVDGVEYHNAKTRPREQVNSTY 285
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 FIFPPKQVLTITLTPKTVCVVDVSDQDPVHFSWFDVDEYHNAQTRPREQVNSTY 179
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 286 RVVSVTVLHODMLNGEYCKVSNKALPAPIEKTISKAGQPREQVYTLPPSRDELTK 345
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 RSVSELPILHODMLNKRTRCKVTSAPSPSIEKISKDEGRTOVPHVYVTSPTKSEMTQ 239
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 346 NQVSLTCLVKGFPSPDIAVWESNGQPENNYKTTTPVLDSDGSFLYSKLTVDKSRWQGG 405
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 240 NEVSTICVWKGFPYPPDIYEWQMGNGQPENNYKNTPTMTDSDSYFLYSKLTVDKSRWQGG 299
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 406 NVFSCSVNHEALHNHYOKSLSLSPG 431
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 NTFCSVLHGLHNHTTKSFSPRTPG 325
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 32

```

S00847
Ig gamma-2c chain C region - rat (fragment)
C/Species: Rattus norvegicus (Norway rat)
C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 23-Jul-1999
C/Accession: S00847
R:Brueggemann, M.; Delmaestro-Galife, P.; Waldmann, H.; Calabi, F.
Eur. J. Immunol. 18, 317-319, 1988
A/Title: Sequence of a rat immunoglobulin gamma-2c heavy chain constant region cDNA: ext
A/Reference number: S00847; MUID:88166903; PMID:3127222
A/Accession: S00847
A/Molecule type: mRNA
A/Residues: 1-329 <BRU>
A/Cross-References: EMBL:X07189; NID:957602; PID:CAA30169.1; PID:g663228
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F/20-84/Domain: immunoglobulin homology <IMM>

```

```

Query Match 34.1%; Score 823; DB 2; Length 329;
Best Local Similarity 58.5%; Pred. No. 4.1e-43;
Matches 151; Conservative 45; Mismatches 56; Indels 6; Gaps 2;

```

```

Qy 179 SGTW-----TCTVLNOKRVEFKIDIVPCAPAPKSCDHTHCPELLGGPSVFLPPPK 233
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 SSTWSSQVTVCSVNAHPATKSNLIKRIEPRPKPRPPDIDSCDNLGGPSVFIPEPK 130
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 234 DTLMTSRPEPVYCVVVDVSHEDPEVKFENNYVDGVEYHNAKTRPREQVNSTYRVVSVLT 293
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 DILMITLTPKTVCVVVDVSEEDPDVQFSWFDVNAVFTAQDPHEQLNGTRFVYSTLHI 190
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 294 LHODMLNGEYCKVSNKALPAPIEKTISKAGQPREQVYTLPPSRDELTKNQVSLTCL 353
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 OHQDMMKSGEFCKKNNKDLPSPIEKTISKPRGKARTQVYTIPPPRQMSKNKYSTLCM 250
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 354 VKGFYSPDIAVWESNGQPENNYKTTTPVLDSDGSFLYSKLTVDKSRWQGNVPSGSVM 413
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 VTSFYPASISVEMENGELEBDYKNTLFLVLDSDSYFLYSKLTVDTSMMRGDITTCVV 310
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 414 HEALHNHYOKSLSLSPG 431
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 311 HEALHNHTTKSLSPG 328
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 33

```

GIMS
Ig gamma-1 chain C region, secreted form - mouse
C/Species: Mus musculus (house mouse)
C/Date: 31-Mar-1980 #sequence_revision 24-Sep-1991 #text_change 16-Jul-1999
C/Accession: A02159; A26234; A26236
R:Hongo, T.; Obata, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takahashi, N.;
Cell 18, 559-568, 1979
A/Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain g
A/Reference number: A02159; MUID:80045036; PMID:115593
A/Accession: A02159
A/Molecule type: DNA
A/Residues: 1-324 <HON>
A/Cross-References: GB:J00453
A/Note: The sequence was determined from the germline gene
R:Obata, M.; Yamawaki-Kataoka, Y.; Takahashi, N.; Kataoka, T.; Shimizu, A.; Mano, Y.; Se
Gene 9, 87-97, 1980
A/Title: Immunoglobulin gamma-1 heavy chain gene: structural gene sequences cloned in a
A/Reference number: A26234; MUID:80202559; PMID:6769752
A/Contents: MOPC 31C
A/Accession: A26234

```

A:Molecule type: mRNA
 A:Residues: 76-324 <GBA>
 A:Cross-references: GB:V00775; NID:951652; PIDN:CAA24153.1; PID:951653
 R:Rogers, J.; Clarke, P.; Salsner, W.
 Nucleic Acids Res. 6, 3305-3321, 1979
 A:Title: Sequence analysis of cloned cDNA encoding part of an immunoglobulin heavy chain
 A:Reference number: A26236; MUID:80012837; PMID:113776
 A:Contents: MOPC 21
 A:Accession: A26236
 A:Molecule type: mRNA
 A:Residues: 170-275, 'D', 277, 'D', 279-322 <ROG>
 A:Cross-references: GB:V00795; NID:951830; PIDN:CAA24176.1; PID:9780265
 R:Adetupbo, K.
 J. Biol. Chem. 253, 6068-6075, 1978
 A:Title: Evolution of immunoglobulin subclasses. Primary structure of a murine myeloma g
 A:Reference number: A26237; MUID:78242288; PMID:68554
 A:Contents: annotation; MOPC 21
 A:Note: This is the final paper in a series reporting the protein sequence, the disulfid
 A:Note: There are a number of differences from the sequence shown
 C:Genetics:
 A:Introns: 1/1; 98/1; 111/1; 218/1
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka
 hain disulfide bonds. In some cases, such as IGA and IGM, the subunits associate into 1a
 C:Superfamily: Immunoglobulin C region; Immunoglobulin homology
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; Immunoglob
 F:20-84/Domain: immunoglobulin homology <IM1>
 F:98-110/Region: hinge
 F:131-200/Domain: immunoglobulin homology <IM2>
 F:237-304/Domain: immunoglobulin homology <IM3>
 F:127-82, 138-198, 244-302/Disulfide bonds: #status experimental
 F:102/Disulfide bonds: interchain (to light chain) #status experimental
 F:104, 107, 109/Disulfide bonds: interchain (to heavy chain) #status experimental
 F:174/Binding site: carbohydrate (Asn) (covalent) #status experimental
 Query Match 33.9%; Score 818.5; DB 1; Length 324;
 Best Local Similarity 51.8%; Pred. No. 7.5e-43;
 Matches 156; Conservative 50; Mismatches 60; Indels 35; Gaps 5;
 QY 133 LLOGSLTLTLESPGSSPSVQCRPRGNIQGGKTLVSQLELDSDGTMCTVLA--NQ 190
 DB 56 VLQSLVLTLSSTVTPSSP-----RPSSTVTCNVNHPASS 90
 QY 191 KKEVEKIDIVPCPAPEPKSCDKTHTCPPELGPSVFLPPPKDKTMTSRPEVTCVVD 250
 DB 91 TKVDKIVPRDGC-KPCIC---TVPEV---SVFIFPPKRVDTLTLPKATCVVD 142
 QY 251 VSHPEPEKENVYGVGVNNAKTPREEQVNSTRVSVLTVLHQMVLNKEKYCKVSN 310
 DB 143 ISKDPPEVQFVFWVDVEVHTAQTQPRREQFNSTRSVSELPIMHQLNLNGEFCRVNS 202
 QY 311 KALPAPIKTSKAGQPREPOVYTLPPSRDELTKNOVSLCGLVGFPPSDIAVWESNG 370
 DB 203 AAFPPPIKTSKTRKPAKAPVYITPPPEKQMAKQVSLCTMTDFPEDTIVEMQNG 262
 QY 371 QPENNYKTPVLDSDGSFFLYSKLTVDKSRWQGNVSCVVMHRLAHNYTOKSLDSP 430
 DB 263 QPAENVKQTQPIMTNGSYFYVSKLVNOKSWNEANFTCTCVLHGLNHHTEKSLNSP 322
 QY 431 G 431
 DB 323 G 323
 RESULT 34
 G2MS11
 19 gamma-2b chain - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 31-Mar-1980 #sequence revision 01-Dec-2000 #text change 01-Dec-2000
 C:Accession: S25057; A02157; A26233; A26233; A53598
 R:Fischer, R.; Voss, A.; Niersbach, M.; Munziker, W.; Hirsch, H.J.; Kreuzaler, F.
 Submitted to the EMBL Data Library, July 1992
 A:Description: Production of a Tobacco mosaic virus (TMV) inactivating neotop specific p
 A:Reference number: S25057

A:Accession: S25057
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-474 <FIS>
 A:Cross-references: EMBL:X67210; NID:954826; PIDN:CAA76649.1; PID:954827
 R:Yamawaki-Kataoka, Y.; Kataoka, T.; Takahashi, N.; Obara, M.; Honjo, T.
 Nature 283, 786-789, 1980
 A:Title: Complete nucleotide sequence of immunoglobulin gamma2b chain gene cloned from m
 A:Reference number: A02157; MUID:80120716; PMID:6765534
 A:Contents: a allele
 A:Accession: A02157
 A:Molecule type: DNA
 A:Residues: 138-161, 'L', 163-189, 'FP', 193-474 <YAM>
 A:Cross-references: GB:J00461
 A:Note: the sequence was determined from the germ-line gene
 R:Tucker, P.W.; Marcu, K.B.; Slightom, J.L.; Blattner, F.R.
 Science 206, 1299-1303, 1979
 A:Title: Structure of the constant and 3' untranslated regions of the murine gamma2b heav
 A:Reference number: A26235; MUID:80081501; PMID:117548
 A:Contents: MPC 11
 A:Accession: A26235
 A:Molecule type: mRNA
 A:Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TU1>
 A:Note: Lys-474 is probably removed posttranslationally
 R:Tucker, P.W.; Marcu, K.B.; Newell, N.; Richards, J.; Blattner, F.R.
 Science 206, 1303-1306, 1979
 A:Title: Sequence of the cloned gene for the constant region of murine gamma2b immunoglob
 A:Reference number: A26232; MUID:80081502; PMID:117549
 A:Accession: A26232
 A:Molecule type: DNA
 A:Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TU2>
 R:Ollo, R.; Rougeon, F.
 Nature 296, 761-763, 1982
 A:Title: Mouse immunoglobulin allotypes: post-duplication divergence of gamma2a and gamma
 A:Reference number: A26233; MUID:82173203; PMID:6803173
 A:Contents: b allele
 A:Accession: A26233
 A:Molecule type: DNA
 A:Residues: 138-161, 'L', 163-189, 'FP', 193-300, 'R', 302-331, 'A', 333-437, 'DI', 440-474 <OLL>
 A:Cross-references: GB:J00461
 R:Kim, H.; Yamaguchi, Y.; Masuda, K.; Matsunaga, C.; Yamamoto, K.; Irimura, T.; Takahash
 J. Biol. Chem. 269, 12345-12350, 1994
 A:Title: O-glycosylation in hinge region of mouse immunoglobulin G2b.
 A:Reference number: A53598; MUID:94216359; PMID:7512967
 A:Accession: A53598
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 234-251 <KIM>
 C:Comment: The a allele sequence is shown.
 C:Genetics:
 A:Introns: 138/1; 236/1; 258/1; 368/1
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa
 hain disulfide bonds. In some cases, such as IGA and IGM, the subunits associate into 1a
 C:Superfamily: Immunoglobulin C region; Immunoglobulin homology
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobu
 F:152-222/Domain: immunoglobulin homology <IM1>
 F:236-257/Region: hinge
 F:281-350/Domain: immunoglobulin homology <IM2>
 F:387-454/Domain: immunoglobulin homology <IM3>
 F:152/Disulfide bonds: interchain (to light chain) #status predicted
 F:164-220, 288-348, 394-452/Disulfide bonds: #status predicted
 F:247, 250, 253, 256/Disulfide bonds: interchain (to heavy chain) #status predicted
 F:324/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 33.9%; Score 818; DB 1; Length 474;
 Best Local Similarity 39.7%; Pred. No. 1.3e-42;
 Matches 184; Conservative 66; Mismatches 124; Indels 90; Gaps 14;
 QY 34 GDTVELCTASQKSIQF--HMKNSTIKTLGNG-----SFL--TKGPSKLNDRADRS 85
 DB 34 GASVNSCKASGYFTFYVWV-----VKQPGQGLEWIGYINPNKGTGKNEKFKGKAT 88
 QY 86 LW--DQGNFPLIITKNKIDSDTYICEVEDQKEEVQLVFGLTANSDTHLQGSILTLT 142

```
Db      89 LTRDKSSNATYMWELSLTSEDSALVYCARD-----YVDYFAVWGCTVLT 134
Qy      143 LSSPSSSSSVGQ-----CRSPGKNIQGG-----KTLVSQGLE 175
Db      135 VSAAKTTPSVYVPLAPGCCDDTTGSSSTSGCLVKGYPEPESVTYWMNSGSLSSSHHTLSQAL 194
Qy      176 LQD-----SGTW-----TCTVLD--NQKKVEKID-----IYPCPAPEPS 209
Db      195 LQSGLYTSSSVTVPSSTPSCQVTGTSVAHPASSTTVDDKLESGGISTINPCP-----P 249
Qy      210 CDKHTTC--PELLGSPVFLPFPKPYDTLMISRTPEYTCVVVDVSHEDPEVKFMYVDGV 267
Db      250 CKECHKCPAPNLEGGSVFIFPPNIDVLMISLTPTKCVVDVSDDDPDVGQISMFVNVN 309
Qy      268 EYHNAKTRPEEQNSTYVVSVLYTLHODMLNGKSKCKVSNKALPAPIEKTISKAKQ 327
Db      310 EYHTAQOTQTHREDYNTIRVSTLPIQHODMWSGKEFKCKVNNKDLPSPIERTISKIKGL 369
Qy      328 PREPOVYTLPPSRDELTKQVSLTCLVKGFPSPDIVEMESNGQPENNYKTPPVLDSDG 387
Db      370 VDAPOVYLLPPPAEQSLSRQDVSLTCLVGFNPEDISVEWTSNGHTEENTKDTAPVLDSDG 429
Qy      388 SFPLYSKLTIVDKSRMOCGVFSCSVMEALHNHYTQKSLSLSPG 431
Db      430 SYFIYSKLMNMTSKMEKTSFSCNVNHEGLKNVYLLKTTISRSPG 473
```

RESULT 35

```
G2MSA
Ig gamma-2a chain C region, secreted form (allele a) - mouse
C.Species: Mus musculus (house mouse)
C.Date: 30-Sep-1980 #sequence_revision 01-Sep-1981 #text_change 16-Jun-2000
C.Accession: A02152; A32657; A32658
R.Sikorav, J.L.; Aulifray, C.; Rougeon, F.
Nucleic Acids Res. 8, 3143-3155, 1980
A.Title: Structure of the constant and 3' untranslated regions of the murine Balb/c gamma
A.Reference number: A02152; MUID:81076554; PMID:677755
A.Accession: A02152
A.Molecule type: mRNA
A.Residues: 1-330 <SIK>
A.Cross-references: GB:V00798; NID:G51835; PIDN:CAA24178.1; PID:G1333984
R.Yamawaki-Kataoka, Y.; Miyata, T.; Honjo, T.
Nucleic Acids Res. 9, 1365-1381, 1981
A.Title: The complete nucleotide sequence of mouse immunoglobulin gamma 2 A gene and env
A.Reference number: A32657; MUID:81198976; PMID:6262729
A.Accession: A32657
A.Molecule type: DNA
A.Residues: 1-330 <YAM>
A.Cross-references: GB:J00470
A.Note: The sequence was determined from the germ-line gene
R.Ollo, R.; Aulifray, C.; Morchamps, C.; Rougeon, F.
Proc. Natl. Acad. Sci. U.S.A. 78, 2442-2446, 1981
A.Title: Comparison of mouse immunoglobulin gamma-2A and gamma-2B chain genes suggests t
A.Reference number: A32658; MUID:81223894; PMID:6787604
A.Accession: A32658
A.Molecule type: DNA
A.Residues: 1-330 <OLL>
A.Note: The sequence was determined from the germ-line gene
R.Bourgois, A.; Fougereau, M.; Rocca-Serra, J.
Eur. J. Biochem. 43, 423-435, 1974
A.Title: Determination of the primary structure of a mouse IgG2a immunoglobulin: amino-a
A.Reference number: A32659; MUID:7417517; PMID:4831970
A.Contents: annotation; myeloma protein MOPC 173
A.Note: this is one paper in a series reporting the sequence; for additional references,
Ride Preval, C.; Fougereau, M.
Eur. J. Biochem. 30, 452-462, 1972
A.Title: Determination of the primary structure of a mouse gammaG2a immunoglobulin. Ider
A.Reference number: A32660; MUID:73056887; PMID:4565406
A.Contents: annotation; MOPC 173, disulfide bonds
C.Genetics:
```

```
A.Introns: 1/1; 98/1; 114/1; 224/1
C.Comment: An immunoglobulin heterotetramer subunit consists of two identical light (kai
hain disulfide bonds. In some cases, such as Iga and IgM, the subunits associate into 1:
C.Superfamily: Immunoglobulin C region; immunoglobulin homology
C.Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob
F:20-84/Domain: immunoglobulin homology <IM1>
F:98-113/Region: hinge
F:137-206/Domain: immunoglobulin homology <IM2>
F:243-310/Domain: immunoglobulin homology <IM3>
F:15/Disulfide bonds: Interchain (to light chain) #status experimental
F:127-82,144-204,250-308/Disulfide bonds: #status experimental
F:107,110,112/Disulfide bonds: Interchain (to heavy chain) #status experimental
F:180/Binding site: carbohydrate (Aen) (covalent) #status predicted
Query Match 33.4%; Score 805.5; DB 1; Length 330;
Best Local Similarity 58.5%; Pred. No. 4,8e-42;
Matches 155; Conservative 33; Mismatches 58; Indels 19; Gaps 5;
```

```
179 SGTW-----TCTVLD--NQKKVEKID-----IYPCPAPEPSCDKHTCPPELLGSPSVF 226
Db      72 SGTWPSQSTTCVVAHPASTKYDKIERGPIKCP--PCKC---PAPVLLGSPSVF 124
Qy      227 LFPKPKDTLMISRTPEYTCVVVDVSHEDPEVKFMYVDGVEVHNAKTRPEEQNSTYR 286
Db      125 IFPPKIKQVLMISLPIVTCVVVDVSDDDPDVGQISMFVNNVEVHTAQOTQTHREDYNSTLR 184
Qy      287 VVSUTVTLHODMLNKEFKCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKN 346
Db      185 VVSALPIQHODMWSGKEFKCKVNNKDLPAPIERTISKRGSVRAPOVYVLPPEPEMTKK 244
Qy      347 QVSLTCLVKGFPSPDIVEMESNGQPENNYKTPPVLDSDGSFPLYSKLTIVDKSRMOCGN 406
Db      245 QVTLTCWYTDMPEDIVYBMTNGKTELYNKTBEVLDSGYSFMYSKLRYEKKMVERN 304
Qy      407 VFSCSVMEALHNHYTQKSLSLSPG 431
Db      305 SVSCGVMEALHNHTTSFSPRTPG 329
```

RESULT 36

```
G2MSA
Ig gamma-2a chain C region, secreted form (allele b) - mouse
C.Species: Mus musculus (house mouse)
C.Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 16-Jul-1999
C.Accession: A02153; A32656
R.Schreier, P.H.; Botwell, A.L.M.; Mueller-Hill, B.; Baltimore, D.
Proc. Natl. Acad. Sci. U.S.A. 78, 4495-4499, 1981
A.Title: Multiple differences between the nucleic acid sequences of the IgG2a(a) and IgG
A.Reference number: A02153; MUID:82037861; PMID:6170065
A.Accession: A02153
A.Molecule type: mRNA
A.Residues: 1-315 <SCH>
A.Cross-references: GB:J00479
A.Experimental source: strain C57BL/6
R.Dognin, M.J.; Laureys, M.; Stroberg, A.D.
Proc. Natl. Acad. Sci. U.S.A. 78, 4031-4035, 1981
A.Title: Multiple amino acid substitutions between murine gamma 2a heavy chain Fc region
A.Reference number: A32656; MUID:82037777; PMID:6794027
A.Accession: A32656
A.Molecule type: protein
A.Residues: 118-267; 'E', 269-328, 'G', 330-334 <DOG>
C.Comment: Lys-335 is removed posttranslationally.
C.Comment: The sequence differs from that of the a allele, from BALB/c mice, at 15 of t
C.Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
hain disulfide bonds. In some cases, such as Iga and IgM, the subunits associate into 1:
C.Superfamily: immunoglobulin C region; immunoglobulin homology
C.Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob
F:20-84/Domain: immunoglobulin homology <IM1>
F:98-118/Region: hinge
F:142-211/Domain: immunoglobulin homology <IM2>
F:248-315/Domain: immunoglobulin homology <IM3>
F:15/Disulfide bonds: Interchain (to light chain) #status predicted
F:127-82,149-209,255-313/Disulfide bonds: #status predicted
```

F108.117/Disulfide bonds: Interchain (to heavy chain) #status predicted
F185/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 33.2%; Score 801; DB 1; Length 335;

Best Local Similarity 50.2%; Pred. No. 9,1e-42;

Matches 166; Conservative 52; Mismatches 77; Indels 36; Gaps 9;

QY 124 GLTASDT--HLAQ---QSLTLTLESPGSS-----PSVQCRSPRGKNIQSGKTLVSQ 173
DB 17 GTTSSVTLGLGVKGFPEPVTLLTNSSGLSSGVHTFPAL-----LQSG-LYLTSS 66
QY 174 LELQDSGTW----TCTV-----LQNKVEFKDIY--PCPADEPKSCDKHTCPPLL 220
DB 67 SVTVSNMWPSTQITCNVAHPASSTKVSKIEPRVITQNPCP---PQORVPCCAPDL 123
QY 221 GSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKFMWYDGVNAKTRPEQ 280
DB 124 GSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPDVQISMFVNNEVHTAQVTHRED 183
QY 281 YNSTRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKQPREPOVYTLPPSR 340
DB 184 YNSTRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKQPREPOVYTLPPSR 243
QY 341 DELTKNQVSLTCLVKGFYPSDIAVEESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKS 400
DB 244 EEMTKKESLTCTMIGFLPAEIAVDMTSGRTEQNKATATVLDSDGSYFMYSKLRVQKS 303
QY 401 RMQGNVSCSVMEHALNHYTKSLSPG 431
DB 304 TWERGSFLPACSVVHEVLNHYTKSLSPG 334

RESULT 37

S01321
Ig gamma-2b chain precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 23-Jul-1999
C/Accession: S01321
R/de Waale, P.; Feys, V.; van de Voorde, A.; Molemans, F.; Fiers, W.
Eur. J. Biochem. 176, 287-295, 1988
A/Title: Expression in non-lymphoid cells of mouse recombinant immunoglobulin directed
A/Reference number: S01320; MUID:88329081; PMID:3138116
A/Accession: S01321
A/Molecule type: mRNA
A/Residues: 1-475 <DE1>
A/Cross-references: EMBL:X13188; NID:G51780; PID:CA31580.1; PID:G51781
A/Note: this sequence was determined from the differentiated gene
C/Superfamily: Immunoglobulin C region; Immunoglobulin homology
C/Keywords: Immunoglobulin
F.1-19/Domain: signal sequence #status predicted <SIG>
F.20-475/Product: Ig gamma-2b chain #status predicted <MAT>
F.159-223/Domain: immunoglobulin homology <IMM>

Query Match 32.9%; Score 793.5; DB 2; Length 475;

Best Local Similarity 40.0%; Pred. No. 3.9e-41;

Matches 187; Conservative 58; Mismatches 135; Indels 87; Gaps 14;

QY 30 LGKSDTVELTCTAS--QKKSIOFHMKNSNOIKILGNG-----SFLTKPSKLNDAAD 81
DB 30 LARPAASVYLSCKASGYTLTSTYISW----VKQRTGGLEWIGELIYSGNSYFNEKFK 84
QY 82 SRSRLM---DQGNFLIILKLIKIEDSDYICEVEDQKEVOLVGLTANSDDLHLOQS 138
DB 85 GKATILTVKSSSTAYLHLSLSLSEDSAVYFCA---GPRQVGLLPRG-----YMQG 132
QY 139 LTLTLESPGSSPSVQ-----CRSPRGKNI-----QSGKTLV--- 171
DB 133 TLVTASAKTTPSPYPLAPGCGDTTGSSTVGLGVKGFPESTVTVWNSGLSSVHTF 192
QY 172 -----SQLELQDSGTW----TCTVQ--NQKKVEFKID-----IYPCAPE 206
DB 193 PALQSGLYTWSSSVTPSTWPSQTVTCVAHPASSTVVDKLEPSPGPTSTINCP--- 249

QY 207 PKSCDKHTTC--PELLGSPVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKFMWY 264
DB 250 --PCKECHKCPAPNLBEGSPVFIFFPNIKQVLMISLTPKTCVVVDVSHEDPDVQISMV 307
QY 265 DGEVHNNAKTRPEEQNSTYRVSUTVLHQMVLNGEKYCKVSNKALPAPIEKTISKA 324
DB 308 NNVEVLTAQVTHREDNSTIRVVSALPIQHODMSKEFKCKVNNKDLPAPIERTISKI 367
QY 325 KQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEESNGQPENNYKTTPVL 384
DB 368 KGIYRQVYTLSPPEQLSRKDVSLTCLAVGSPEDISVEMTSGNTEENYKDTAPVLD 427
QY 385 SDGSFFLYSKLTVDKSRMQGNVSCSVMEHALNHYTKSLSPG 431
DB 428 SDGSFFLYSKLTVDKSRMQGNVSCSVMEHALNHYTKSLSPG 474

RESULT 38

PS0019
Ig gamma-2a chain C region - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 16-Jul-1999
C/Accession: PS0019; D25941
R/Bruggemann, M.
Gene 74, 473-482, 1988
A/Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.
A/Reference number: PS0017; MUID:89232738; PMID:3319946
A/Accession: PS0019
A/Molecule type: DNA
A/Residues: 1-322 <BRU>
R/Bruggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Walimann, H.
Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986
A/Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody
A/Reference number: A25941; MUID:86287397; PMID:3016742
A/Accession: D25941
A/Molecule type: DNA
A/Residues: 216-322 <BR2>
C/Genetics:
A/Introns: 98/1; 109/1; 216/1
C/Superfamily: Immunoglobulin C region; immunoglobulin homology
C/Keywords: Immunoglobulin
F.20-84/Domain: immunoglobulin homology <IMM>

Query Match 32.8%; Score 792.5; DB 2; Length 322;

Best Local Similarity 51.0%; Pred. No. 2.9e-41;

Matches 160; Conservative 47; Mismatches 72; Indels 35; Gaps 8;

QY 133 LLAG---QSLTLTLESPGSS-----PSVQCRSPRGKNIQSGKTLVSQLELQDSGTW-- 182
DB 28 LVKGFPEPVTYVWNSGALSSGVHTFPAY-----LQSGLYTLNSSLTV--PSSTWSS 77
QY 183 ---TCTVQ--NQKKVEFKDIYPCPADEPKSCDKHTCPCELLAGSPVFLPPPKKDTLM 237
DB 78 QAVTCVNAHPASSTKVSKIEPRVITQNPCP---NPGCTGSGV-----SSVTFPPPKKDTLM 127
QY 238 ISRTPEVTCVVVDVSHEDPEVKFMWYDGVNAKTRPEEQNSTYRVSUTVLHOD 297
DB 128 ILLTPKTCVVVDVSHEDPEVKFMWYDGVNAKTRPEEQNSTYRVSUTVLHOD 187
QY 298 WLNGEKYCKVSNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGF 357
DB 188 WLNGEKYCKVSNKALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGF 247
QY 358 YPSDIAVEESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRMQGNVSCSVMEHAL 417
DB 248 YPDIYTEMONGPQGNKYNTPTMTDSDSYFLYSKLTVDKSRMQGNVSCSVMEHAL 307
QY 418 HNHYTOKSLSPG 431
DB 308 HNHYTEKSLSPG 321

F:322-372/Domain: immunoglobulin homology <IMM>

Query Match 24.9%; Score 601.5; DB 2; Length 459;

Best Local Similarity 41.9%; Pred. No. 1.9e-29;

Matches 153; Conservative 47; Mismatches 98; Indels 67; Gaps 11;

QY 1 MNRGVPFRHLVLVQLALLPAATQGNKYLGGKGDVETLTASQKSIQPHMKNSNOIK 60

DB 1 MNRRIYFQCLLVLPALLPAATGWKTVRGKAGIVELPCQSSQKRNVSFNMKXANQVK 60

QY 61 ILNGG---SFLTGPKSLNDRADRSRSLMDQGNFPLIKNLKIEDSDTYICEVDQKE 116

DB 61 ILNGGSSSSSSFWLGNPSLNRKVSXKMMDQGFPLVIXDLRMDDGTYICEVDQK 120

QY 117 EVQLVFGITANSDPHLLQGSLTTLTSPSSPVQCRSPRGKNIQGGKTLVSQLEL 176

DB 121 EVELLVFRLTANPNRLHGLSLTLTLEGPSVGSVQWKSPENKIETGFLCSNPKRL 180

QY 177 QDSGTWCTV-LQNGKVEFKIDIVCPAPEKSCDKHTHCPBELLGSPVLPFPKPKDT 235

DB 181 QDSGTWCSGLHSFQDNKLELDIKIIVLGFPKASA-----TYVKKGEQVFSFP----- 229

QY 236 LMISTPEVTCVVDVSHEDPEV--KFNWYVDGVVHNAKTPREEOYNSTRVVSULTV 293

DB 230 -----LNFEDESLGELMWQVDGAS----- 249

QY 294 LHQDVLNGKVEKCKVS-NKALPAPIEKTISKAKGPREPOVYTLPPSRDELTL-KNQVSLT 351

DB 250 SAQSVVSSLEBRKVSQKILP---DLKIQMSKGLPLS---LTLPLALHRYAGSGNLSLT 303

QY 352 CLVKG 356

DB 304 -LDKG 307

RESULT 43

A36040

IG heavy chain V-III region (ART) - human (fragments)

C:Species: Homo sapiens (man)

C>Date: 16-Nov-1990 #sequence_revision 16-Nov-1990 #text_change 21-Jan-2000

C:Accession: A36040

R:Leitz, M.; Weiss, D.T.; Solomon, A.

Proc. Natl. Acad. Sci. U.S.A. 87, 6542-6546, 1990

A:Title: Immunoglobulin heavy-chain-associated amyloidosis.

A:Reference number: A36040; MUID:90370821; PMID:2118650

A:Accession: A36040

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-218 <EUL>

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:132-199/Domain: immunoglobulin homology <IMM>

Query Match 23.8%; Score 574.5; DB 2; Length 218;

Best Local Similarity 77.5%; Pred. No. 3.5e-28;

Matches 110; Conservative 3; Mismatches 12; Indels 17; Gaps 1;

QY 250 VLTVHOMLNGKEYKCKVSKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKQVS 349

DB 94 VAAALLTMDGQ-----TLVTSSSGPREPQVTLPPSRDELTKQVS 136

QY 350 LTCLVKGFPSPDIAMWESNGOPENNYKTPPVLDSDGSFLLYSKLTVDKSRMOQGNFS 409

DB 137 LTCLVKGFPSPDIAMWESNGOPENNYKTPPVLDSDGSFLLYSKLTVDKSRMOQGNFS 196

QY 410 CSVMHEALHNYTKQSLSPG 431

DB 197 CSVMHEGLHNYTKQSLSPG 218

RESULT 44

S30193

T-cell surface glycoprotein CD4 - dog

C:Species: Canis lupus familiaris (dog)

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000

C:Accession: S30193

R:Milde, K.F.; Conner, G.E.; Mintz, D.H.; Alejandro, R.

Biochim. Biophys. Acta 1172, 315-318, 1993

A:Title: Primary structure of the canine CD4 antigen.

A:Reference number: S30193; MUID:93192324; PMID:7916632

A:Accession: S30193

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-432 <MTL>

C:Cross-references: EMBL:X68565; NID:9288652; PID:CAB37664.1; PID:94467377

C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

C:Keywords: glycoprotein

F:202-311/Domain: immunoglobulin homology <IMM>

Query Match 23.8%; Score 574.5; DB 2; Length 432;

Best Local Similarity 34.0%; Pred. No. 7.7e-28;

Matches 155; Conservative 61; Mismatches 127; Indels 113; Gaps 15;

QY 12 LVQLALLPAATQGNKYLGGKGDVETLTASQKSIQPHMKNSNOIKILNGGSPFLTK 71

DB 1 LMLQVNLPAVTPREVVLGKAGDAVELPCOTSQKNIHNRDSSNVQILGNQGSFTV 60

QY 72 GPSRLNDRADRSRSLMDQGNFPLIKNLKIEDSDTYICEVDQKEVQLVFGTLA----- 127

DB 61 GSSRLKGRVSKKSLMDQGFPLVIXDLRVADSGITFCDF--DKRQEVLLVFNLTAKWDS 119

QY 128 -----NSDTHLQGSITLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTW 182

DB 120 GSSSSSNIHLQGOQLTLTLNPSGSSPSVQWGPKNKXHGQNLSTLSPBELQDSGTW 179

QY 183 TCTVLQNGKVEFKIDIVCPAPEKSCDKHTHCPBELLGSPVLPFPKPKDTLMISRT 242

DB 180 TCIISQSKTVERINVLVAF--QVSNTPFAKE--GQVFEFSP----- 221

QY 243 EYTCVVDVSHEDPEV--KFNWYVDGVVHNAKTPREEOYNSTRVVSULTV 300

DB 222 -----LSFEDENLVGELRMQAGAS-----SSLL-----WIS 246

QY 301 GKVEKCKVSKALPAPIEKTISKAKGPRE--POVYTLPPSRDELTKQVSITLVYGFY 358

DB 249 FTLENRLSKMEKAAAP-----KLQMSKSLPLRFLPVLSRYAGSGITLTLAKGTL 301

QY 359 PSDIAMEWESNGOPENNYKTPPVLDSDGSFLLYSKLTVDKSRMOQGNVSVMHEALH 418

DB 302 YQEV-----NLVVRANSQNNL--TCEVLGP--- 326

QY 419 NHYTKSLSPGILDETCAEADGELDGLMTDP 454

DB 327 ---TSPELTIS--LNKEGAQAKSK-QQKLVWVVDP 356

RESULT 45

S69340

IG heavy chain VHIII-D-JH-CH3 region - human

C:Species: Homo sapiens (man)

C>Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 21-Jan-2000

C:Accession: S69340

R:Kamlich, A.A.; Aucouturier, P.; Preud'homme, J.L.; Cogne, M.

Eur. J. Biochem. 229, 54-60, 1995

A:Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.

A:Reference number: S69339; MUID:95265687; PMID:7744049

A:Accession: S69340

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-249 <KHA>

A:Cross-references: EMBL:X81696

C:Superfamily: immunoglobulin C region; immunoglobulin homology

F:162-229/Domain: immunoglobulin homology <IMM>

Query Match

Best Local Similarity 23.7%; Score 572.5; DB 2; Length 249;

Matches 155; Conservative 62.9%; Pred. No. 5.4e-28;

Matches 117; Conservative 9; Mismatches 23; Indels 37; Gaps 3;
QY 266 GVEVNAKTKPREEOYNSRYRVSVLT-----VLHODMLNGEKYCKKVN 310
DB 80 GDSVGRFTISRDNSENQILQWNSLTAEDTAAYCAKVGSVYH----- 124
QY 311 KALPAPIE-----KTKSKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAB 365
DB 125 --LPSAIHWMGGOTRIYVSGQPREPOVYTLPPSRDEMTKNQVSLTCLVKGFPSPDIAB 182
QY 366 MESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFGSCVMEALHNHYTKS 425
DB 183 MESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFGSCVMEALHNHYTKS 242
QY 426 LSLSPG 431
DB 243 LSLSPG 248
RESULT 46
S14236
Ig gamma-1 chain C region (15C5) - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 23-Jul-1999
C/Accession: S14236
R/Vandamme, A.M.; Bulens, F.; Bernar, H.; Nelles, L.; Lijnen, R.H.; Collen, D.
Eur. J. Biochem. 192, 767-775, 1990
A/Title: Construction and characterization of a recombinant murine monoclonal antibody c
A/Reference number: S14236; MUID:91006173; PMID:2209622
A/Accession: S14236
A/Molecule type: mRNA
A/Residues: 1-152 <VANA>
A/Cross-references: EMBL:56393; NID:951617; PIDN:CAA39804.1; PID:951618
C/Superfamily: Immunoglobulin C region; Immunoglobulin homology
C/Keywords: Immunoglobulin
F/65-132/Domain: Immunoglobulin homology <IMM>
Query Match 23.4%; Score 566; DB 2; Length 152;
Best Local Similarity 61.6%; Pred. No. 7.6e-28;
Matches 93; Conservative 33; Mismatches 25; Indels 0; Gaps 0;
QY 281 YNSTYRVSVLTGLVHODMLNGEKYCKKVSNAKLPAPIEKTKISKAKGQPREPOVYTLPPSR 340
DB 1 FSTSPSVSELPIHQMDMLNGEKYCKKRVNSAFPAIEKTKISKAKGQPREPOVYTLPPSR 60
QY 341 DELTKNOVSLTCLVKGFPSPDIABMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKS 400
DB 61 EQMADKVSILTCTMTDFPFEDITVEWQNGQPAENYKNTQPIMTNTGYSFYVSKLNVQKS 120
QY 401 RMOGQNVFGSCVMEALHNHYTKSLSLSPG 431
DB 121 NMEAGNTFTCSVLHESGLHNHHTKSLSHSPG 151
RESULT 47
A27449
T-cell surface glycoprotein CD4 precursor - rat
N/Alternate names: W3/25 antigen
C/Species: Rattus norvegicus (Norway rat)
C/Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 21-Jan-2000
C/Accession: A27449; A35433
R/Clark, S.J.; Jeffries, W.A.; Barclay, A.N.; Gagnon, J.; Williams, A.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 1649-1653, 1987
A/Title: Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for deriv
A/Reference number: A27449; MUID:87175535; PMID:3104900
A/Accession: A27449
A/Molecule type: mRNA
A/Residues: 1-457 <CTAA>
A/Cross-references: GB:M15768; NID:9203387; PIDN:AAA0901.1; PID:9203388
R/Davies, S.U.; Ward, H.A.; Puklavec, M.J.; Willis, A.C.; Williams, A.F.; Barclay, A.N.
J. Biol. Chem. 265, 10410-10418, 1990
A/Title: High level expression in Chinese hamster ovary cells of soluble forms of CD4 T
A/Reference number: A35433; MUID:90285164; PMID:2113054

A/Contents: annotation
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C/Keywords: glycoprotein; membrane protein; surface antigen
F/219-300/Domain: Immunoglobulin homology <IMM>
Query Match 20.5%; Score 495; DB 2; Length 457;
Best Local Similarity 32.8%; Pred. No. 5.7e-23;
Matches 137; Conservative 57; Mismatches 133; Indels 92; Gaps 14;
QY 1 MNRGVPPFHL--LVVLQALLPAAIQGNKVVYLGKGGTVELCTASQKKSIOFHKNSNQ 58
DB 1 MCRGFSFPHLLPFLILQLSLKLVVYQGTVVYLGKGGSAELPCESTRSRSSAFWAKSDQ 60
QY 59 IKILNGQSFLLTKGSKINDRADRSRLMDQGNFLIKNLKIEDSDPYICEVEOQKEV 118
DB 61 KTLIGYKKNLLIKGSLKLEYSRDSKKNMERGSPFLINKLMEDSQTYVCELEKKEEV 120
QY 119 QLVFGLTANSPDTHLQGSQSLTLTLES--PGSSPSVQCRSPRGKNIQCGKTLVSQLEIQ 177
DB 121 ELWVRVATVFNQTRLLQGSQSLTLIDSNPXYSDPIECKKSSNVKDSKAFSTSLAIQ 180
QY 178 DSGTWTCTVLONQKVEF--KIDIVCPAPBPCKSCDKTHTCEPILGSPVFLPPKPYDT 235
DB 181 DSGINWCTVTLNQKGSFDMKLSVL-----GFASTSITAYKSGESAEFSFP----- 227
QY 236 LMISTPEVTCVVDVSHDEPEVKRWYVDGVEVNAKTKPREEOYNSRYRVSVLTVLH 295
DB 228 -----LNLGESSLQGLRW-----KAERKASS----- 249
QY 296 QDMLNKGEYCKKVSNAKLPAPIEKTKISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVK 355
DB 250 QSMIFSLKQKVS-----VQKSTSPKQLSE--TLP-----LTLQI- 285
QY 356 GFYPSDIABMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFGSCVM 413
DB 286 ----PQVSLQFAGSG-----NLTLT---HDR-GILYQEVNLVVMKVTQPSNTLTCEVM 331
RESULT 48
RNMST4
T-cell surface glycoprotein CD4 precursor - mouse
N/Alternate names: T-cell differentiation antigen L3T4; T-cell surface antigen T4/Leu 3
C/Species: Mus musculus (house mouse)
C/Date: 30-Jun-1987 #sequence_revision 30-Jun-1997 #text_change 16-Jul-1999
C/Accession: A02110; A26038; A39893; A39855; I54564; I59018; A47642
R/Tourville, B.; Gorman, S.D.; Field, E.H.; HunnKapiller, T.; Parne, J.R.
Science 234, 610-614, 1986
A/Title: Isolation and sequence of L3T4 complementary DNA clones: expression in T cells
A/Reference number: A02110; MUID:87018845; PMID:3094146
A/Accession: A02110
A/Molecule type: mRNA
A/Residues: 1-457 <TQD>
A/Cross-references: GB:M13816; NID:9192070; PIDN:AAA37267.1; PID:9309112
R/Litman, D.R.; Gettner, S.N.
Nature 325, 453-455, 1987
A/Title: Unusual insertion in the immunoglobulin domain of the newly isolated murine CD4 (L
A/Reference number: A26038; MUID:87115821; PMID:3027575
A/Accession: A26038
A/Molecule type: mRNA
A/Residues: 1-457 <LIR>
A/Cross-references: GB:X04836; NID:950353; PIDN:CAA28559.1; PID:950354
R/Gorman, S.D.; Tourville, B.; Parne, J.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 7644-7648, 1987
A/Title: Structure of the mouse gene encoding CD4 and an unusual transcript in brain.
A/Reference number: A39893; MUID:88041159; PMID:2823269
A/Accession: A39893
A/Molecule type: DNA
A/Residues: 1-25, 'E', 27-457 <GGR>
A/Cross-references: GB:M17080; GB:J03003; NID:9192515; PIDN:AAA37402.1; PID:9387124
R/Maddon, P.J.; Molinaux, S.M.; Maddon, D.E.; Zimmerman, K.A.; Godfrey, W.; Alt, F.W.;
Proc. Natl. Acad. Sci. U.S.A. 84, 9155-9159, 1987
A/Title: Structure and expression of the human and mouse T4 genes.
A/Reference number: A39855; MUID:88097446; PMID:3501122

```

A:Accession: A39955
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 25-457 <MAD>
A>Note: the cited Genbank accession number, J03564, is not in release 101.0
R:Partner: J.R.; Hunzapper, T.
Immunol. Rev. 100, 109-127, 1987
A:Title: L374 and the immunoglobulin gene superfamily: New relationships between the im
A:Reference number: 154564; MUID:86152875; PMID:3326818
A:Accession: 154564
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-457 <RES>
A:Cross-references: GB:M36850; NID:9198670; PIDN:AAA39401.1; PID:9198671
A:Accession: 169018
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 208-318 <RE2>
A:Cross-references: GB:M36851; NID:9198672; PIDN:AAA39402.1; PID:9554183
R:Classton, B.J.; Tsagaratos, J.; Kitzbaum, L.; Maddox, J.; Mackay, C.R.; Brandon, M.; M
Immunogenetics 23, 129-132, 1986
A:Title: The L374 antigen in mouse and the sheep equivalent are immunoglobulin-like.
A:Reference number: A47642; MUID:8616694; PMID:3082751
A:Accession: A47642
A:Molecule type: protein
A:Residues: 27-43 <CLA>
C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells t
C:Genetics:
A:Introns: 18/1; 74/1; 128/1; 207/1; 319/1; 386/1; 425/3; 448/2
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: alternative initiators; duplication; glycoprotein; T-cell; transmembrane pro
F:1-26/Domain: signal sequence #status predicted <SIG>
F:127-457/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>
F:135-114/Domain: immunoglobulin homology <IM1>
F:139-189/Domain: immunoglobulin homology #status atypical <IM2>
F:220-301/Domain: immunoglobulin homology <IM3>
F:241-457/Product: CD4, brain-specific short form #status predicted <BRA>
F:321-372/Domain: immunoglobulin homology <IM4>
F:395-419/Domain: transmembrane #status predicted <TM>
F:420-457/Domain: intracellular #status predicted <INT>
F:442-112,158-188,328-370/Disulfide bonds: #status predicted
F:187,296,323,392/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.7%; Score 475; DB 1; Length 457;
Best Local Similarity 52.8%; Pred. No. 9.5e-22;
Matches 104; Conservative 32; Mismatches 55; Indels 6; Gaps 5;

QY 1 MNRGVPPRH-LILVQLALLPAATGKRVGKGDVLTCTASQKKSIOFHKNSQI 59
DB 1 MCRALSLRLLLLLQLQLAVLTGKTLVKGESAEALPCSSQKKTFTVTFWFSQD 60
QY 60 KILGQG-SPLTKG--PSKLNDRADSRSLMDQGFPLINKLKEDSDTYICEVEDKE 116
DB 61 KILGQHGKGVLTIRGSSPSQF-DRFDISKAGWEKFFPLINKLKEDSDTYICELENNKE 119
QY 117 EVQLVLPGLTANSDTHLQGSLLTLLES-PPGSSPSVQCRSPRKNIOGKTLVSQLE 175
DB 120 EVELVVFVKYTPSPGSLIQGSLTLTLDNSKVSNPRLTECKHKGKGVVSGKVLMSMLR 179
QY 176 LODSGTWTCTVQONOK 192
DB 180 VQSDSPFNCTVLTLDOK 196

RESULT 49
Ig Y heavy chain (7.8S) - duck
N:Alternate names: Ig gamma chain (7.8S)
C:Species: Anas platyrhynchos (domestic duck)
C>Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: B46529; S20759
R:Magor, K.E.; Marr, G.W.; Middleton, D.; Wilson, M.R.; Higgins, D.A.
J. Immunol. 149, 2627-2633, 1992

```

```

A:Title: Structural relationship between the two IgY of the duck, Anas platyrhynchos: mcJ
A:Reference number: A46529; MUID:93017865; PMID:1401901
A:Accession: B46529
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-572 <MAG>
A:Cross-references: EMBL:X65219; NID:962442; PIDN:CAA46322.1; PID:962443
A:Experimental source: spleen
A>Note: sequence extracted from NCBI backbone (NCBI:P116127)
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:37-120/Domain: immunoglobulin homology <IMM>

Query Match 17.0%; Score 411.5; DB 2; Length 572;
Best Local Similarity 30.0%; Pred. No. 9.1e-18;
Matches 103; Conservative 55; Mismatches 136; Indels 49; Gaps 10;

QY 102 ESDPTVCEV--EDQKEVQLVFGLTANSDTHLQGSLLTLSPSPSSPSVQCRSP 158
DB 265 DDSVELLCVITGSPSPVEVEMLVDAFAH-----LVATMTRPQ-----REA 306
QY 159 RGNKIQGKTLVSQLELDQSGTWTCTVLDQ-----NKKVEPKIDIVPCPAPKSCDK 212
DB 307 GSKTYMATSGTNVSRBWKGAKFTCKVKHAPATGTAQGARF-----CPGSGAQSCSP 360
QY 213 THTCPELLGSPVFLPPEPKDITLMSRTEPVTCVVVDVSHEDPEVFNNYVDGEVHNA 272
DB 361 I-----QIFVVPSP--GSLYIRQDAKHCLVNLN-SDASLSISWTRKSGALRP 408
QY 273 KTKREKQVNSTYRNVSVLTVLHODWLNKGEYCKRKNKALPAIEKTISKAKQPREPQ 332
DB 409 DPWLTTHFGTGFVASSSLAISTQDMAGRFCTVQHEDLPVPLGSIKHAQKTAAPY 468
QY 333 VYTLPSRDELTKQVSLTCLVKGFPSPDIAVEW--ESNQPENNYKTPPVLN--SDGS 388
DB 469 IFTFPPAEISLAELVTLVKGFPDPEHVEVQMLRNHNSVPAAEFTTTPLEKPNQDGT 528
QY 389 FFYLSKLTVDKSRVQGNVFCGVMHEALHNHYTKSLSPG 431
DB 529 FFLSKKTVTKASQGWGVSAQVHAGLPMRFRGLQKTPG 571

RESULT 50
S04845
Ig heavy chain precursor - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 26-Aug-1999
C:Accession: S04845; S05695
R:Amemiya, C.T.; Haire, R.N.; Liltman, G.W.
Nucleic Acids Res. 17, 5388, 1989
A:Title: Nucleotide sequence of a cDNA encoding a third distinct Xenopus immunoglobulin
A:Reference number: S04845; MUID:89345103; PMID:2503814
A:Accession: S04845
A:Molecule type: mRNA
A:Residues: 1-549 <AME>
A:Cross-references: EMBL:X15114
R:Liltman, G.W.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05695
A:Accession: S05695
A:Molecule type: mRNA
A:Residues: 'LC', 3-308, 'H', 310-549 <LIT>
A:Cross-references: EMBL:X15114; NID:964799; PID:9763031
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: glycoprotein; heterotetramer; immunoglobulin
F:26-109/Domain: immunoglobulin homology <IMM>
F:281,294/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 16.7%; Score 402.5; DB 2; Length 549;
Best Local Similarity 29.1%; Pred. No. 3.1e-17;
Matches 118; Conservative 65; Mismatches 134; Indels 89; Gaps 17;

QY 87 WDQG-----NFFLIK-----NLKIEDSD-----TYICEVE----- 112

```

```
Db 171 MNGSITSGKPNPAVLQSGGLFASSQSLTIPLSDMKAKKSPFCNVEHKTSTKTYOKIE 230
113 --DQKEVQ--LLVGLTANSPT-----HLOQSGITLTLESPPGS 149
231 CODEPPIETVEILOGPCASSKSVELLCLITGAPSEIKVHMLLNGQ--VTNISPSNS 287
150 SPVQCRSPRGKNIQGGKTLVSQLELOSGTWTCTVLOKQKVEPKRIDIVPAPAEPS 209
288 KP---CKENG-TPSSRSKSVKPKEDNSPDSITCKVTH-----PASHTKT 329
210 CDKHTCPPELLGSPVFLPPPKEDTLMTISRTPEVTCVVVDVSHEDPEVKFM--YVDGYE 268
330 EASTKCKDETAIIPKVDVLPSPKID--LVTKKAKVCVLSRMASTD--DLTVQMSRSDGKK 387
269 VHNNAKTPREBOYNSTYRVVSVLTVLHQMVLNGEKYCKVSKALPAPIEKTISKAKGP 328
388 ALAFDSAP--EKAYDGTFTVVKSTLKISPGDWENKQFCNCKVHPDLSPLEKSIQKQDPG 446
329 REPOVYVTLPPSRDELTKNOVSLTCLVKGFPSPDIAMWESNG--QENNYKTTPTPLDSD 386
447 TEFITLTPSPDDELKNDPISLCLMKNPFPDIIYVFMKKGDTLEEDYMTTTPVLEEE 506
387 GSFFL-YSKLTVDKSRWQGNVSCSVMEHALNHNHYTOKSLSLSPG 431
507 EBSGIFSFSLTIARSDMKMGATYSCI-----AAHNTISQNDIKKNGR 548
```

RESULT 51

```
S14683
Ig mu chain precursor, membrane-bound (clone 201) - human
C|Species: Homo sapiens (man)
C|Date: 31-Dec-1991 #sequence, revision 31-Dec-1991 #text_change 23-Jul-1999
C|Accession: S14683; S08047
C|Friedlander, R.M.; NusseZweig, M.C.; Leder, P.
Nucleic Acids Res. 18, 4278, 1990
A|Title: Complete nucleotide sequence of the membrane form of the human IgM heavy chain.
A|Reference number: S14683; MID:90332450; PMID:2115996
A|Accession: S14683
A|Molecule type: mRNA
A|Residues: 1-627 <FRI>
A|Cross-references: EMBL:X17115; NID:G33450; PIDN:CAA34971.1; PID:G33451
C|Superfamily: immunoglobulin C region; immunoglobulin homology
C|Keywords: immunoglobulin; membrane protein
F|1-15/Domain: signal sequence #status predicted <SIG>
F|16-527/Product: Ig mu chain #status predicted <MAT>
F|34-117/Domain: immunoglobulin homology <IMM>
```

Query Match 16.1%; Score 389; DB 2; Length 627;

Best Local Similarity 23.9%; Pred. No. 2.4e-16; Matches 116; Conservative 77; Mismatches 167; Indels 126; Gaps 16;

```
36 TWELTCTASQ--KKSIOFHWKNSNQIKILGNQGSFLTQGPSKLNDRADSRSLMDQGNP 93
175 SVAVGCLADPLPDSITFSWKYKNSDISSTRG-----FP 209
94 LIITKLKIEDSDTYICEVEDQKEVQLVFGLTANSDTLL-----QGOSLTLT 142
210 SVLRGKRYATSOVLPLPSKD-----VMQGDDEHVACVQHPNGNKEKNVPLPI 258
143 LESPPGSSPSVQCR-----SPR-----GKNIQGKTLISVQ 173
259 AELPPVAVSVFVPRDGFPGNPNRSKSLICQATGFSFRQIOVSMKREGKQVSGVTTDQVQ 318
174 LELQDSG-----TWTCVLONQKVEFKLIDIVCPAPEKSCDK 212
319 AEAKEGSPPTYKVTSTLTIKESDMLSQSMFTGCV--DHRGLTFQGNASSMCVPDQTAIR 376
213 THTCPBLLGSPSVFLPPPKKDTLMTISRTPEVTCVVVDVSHEDPEVKFMWYVDGYEYHNA 272
377 -----VFAIPPS-FASIFLTKSTKLCLVTLDTLTYD-SVTISIRONGEAVKT 422
273 KTKPREQYNSTYRVVSVLTVLHQMVLNGEKYCKVSKALPAPIEKTISKAKGP-REP 331
```

```
Db 423 HNNISESHNATFPAVAGEASICEDDMNGERTCTVTHTDLPSPKQTIISRKGVALLHRP 482
332 QVYTLPPSPDELT-KNOVSLTCLVKGFPSPDIAMWESNGQP--ENNYKTTPTPLD--SD 386
483 DYVLLPAPAEQNLRESATITLTCTVGFSPADVFVQMGQGLSPSEKYTASAPMBEPQAP 542
387 GSFFLXSLTVDKSRWQGNVSCSVMEHALNHNHYTOKSLSLSPGLQDETCAEQDGL 446
543 GRYPHSHILTYSSEEMNTGETYTCVAHAPALPNRTERTVDS-----TEGEVSDDEEG 597
447 DGLMTT 452
598 ENLMAT 603
```

RESULT 52

EHMS

Ig epsilon chain C region (version 1) - mouse (fragment)

C|Species: Mus musculus (house mouse)

C|Date: 13-Jun-1983 #sequence, revision 13-Jun-1983 #text_change 16-Jul-1999

C|Accession: A02144

R|Liu, F.T.; Albrandt, K.; Sutcliffe, J.G.; Katz, D.H.

Proc. Natl. Acad. Sci. U.S.A. 79, 7852-7856, 1982

A|Title: Cloning and nucleotide sequence of mouse immunoglobulin epsilon chain cDNA.

A|Reference number: A02144; MID:83117774; PMID:6818553

A|Accession: A02144

A|Molecule type: mRNA

A|Residues: 1-388 <LIU>

A|Cross-references: GB:J00476; NID:9194875; PIDN:AAA38085.1; PID:9387220

C|Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap

ha) disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la

C|Superfamily: immunoglobulin C region; immunoglobulin homology

C|Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin

F|1-44/Domain: immunoglobulin homology <IM2>

F|81-149/Domain: immunoglobulin homology <IM3>

F|186-254/Domain: immunoglobulin homology <IM4>

F|290-361/Domain: immunoglobulin homology <IM5>

F|10,51,62,133,205,228,332,382/Binding site: carbohydrate (Asn) (covalent) #status predi

Query Match 16.1%; Score 388; DB 1; Length 388;

Best Local Similarity 28.3%; Pred. No. 1.6e-16; Matches 119; Conservative 77; Mismatches 138; Indels 86; Gaps 20;

```
38 ELTCTASQKSTIQFHWKNSNQIKILGNQGSFLTQGPSKLNDRADSRSLMDQGNPFLIK 97
22 ELKVTTSQVTS---WGKSAK-----NFTCHVTHPPSFNIESRT-----ILVR 59
98 NLKIEDSDTYI---CEVEDQKEVQL--LVFGLTAN--SPTHLLOQSGSLTLTLESPPGS 149
60 PNATIEPTLELHSSCDNPAHFSTIQLCFYIGHLNDVSVSWLMDREITDTL----- 113
150 SPVQCRSPRGKNIQGGKTLVSQLELOSGTWTCTVLOKQKVEFKLIDIVCPAPEKPS 209
114 AQTVALKE--EGLASTCKKNTTBQOMSESTFTKV--TSGQVDVLAHTRCPDHER-- 169
210 CDKHTCPBLLGSPSVFLPPPKKDTLMTISRTPEVTCVVVDV--SHEDPEVKFN----- 261
170 -----GVITYILPSPILD-LYQNGAPKLCLLVLDLSEKNNVNTVMNOEKRTSV 216
262 ----WYVDGYEYHNAKTPREBOYNSTYRVVSVLTVLHQMVLNGEKYCKVSKALPAPI 317
217 SASQNY---TKHN-----NATTSITSLPVAKMDIEGYQCIYDHPDPPPI 263
318 EKTISKAGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAMWESNGQENN-- 375
264 VRSITKTEGQGSAPRVYVFPPEEB--SDKKTITLTCLTNFPPDISVQMLDGLKILNSQ 322
376 YKTTPTPLDSDGS---FLYSKLTVDKSRWQGNVSCSVMEHALN--HYTOKSLSLSPG 431
323 HSTTP-LKNSGNSNGFPFIFRLKVAKTLTORQFTQOVTHALQKPRKLEKTIISTLG 381
```

RESULT 53
S15590
Ig heavy chain - human
C:Species: Homo sapiens (man)
C:Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S15590
R:Neale, G.A.M.; Kitchingman, G.R.
Nucleic Acids Res. 19, 2427-2433, 1991
A:Title: mRNA transcripts initiating within the human immunoglobulin mu heavy chain enhancer
A:Reference number: S15590; MUID:91252286; PMID:1904154
A:Accession: S15590
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-474 <NEA>
A:Cross-references: EMBL:X58529
A>Note: The authors translated the codon CAA for residue 265 as Glu
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:344-415/Domain: immunoglobulin homology <IMM>

Query Match 16.0%; Score 386.5; DB 2; Length 474;
Best Local Similarity 23.7%; Pred. No. 2.4e-16;
Matches 115; Conservative 78; Mismatches 167; Indels 125; Gaps 16;

```

QY 36 TWELTCTASQ--KKSIOFHMKNSNOIKILNQGSLFKGPSKLNDRADSRSLMDQGNFP 93
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 23 SVAVGCLAQDFLPDSITFSWKYKNSDISSTRG-----FP 57

QY 94 LIINKLKIEDSDTYCEVEDQKEVQLLVFGLTANSDFHL-----QGSLTLT 142
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 58 SVLRGKTAATSQVLPSKDV-----VMQGTDEHVCKVQHPGNKEKNVPLRVI 106

QY 143 LSPSPGSSPSVQCR-----SPR-----GKNIOGKTLVSQ 174
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 AELPKVSVFVPRDGFPGNPKRSKILCOATGFSPRQIQVSWLRGKQVSGVTTDQYA 166

QY 175 ELQDSG-----TWTCVLOQOKVEFKIDIVCPAPBPSCKT 213
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 167 EAKESGPTTYKVTSLITIKESDMLSQSMFTCRV--DHRGLTFQGNASSMCPDDDTAIR- 223

QY 214 HTCPBLGGPSVFLPPPKKDTLMSRPEVTCVAVDVSHEDPEKFMVYDVGVNNAK 273
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 224 -----VFALPPS-FASIFLTSTKLCLVTLDTLTYD-SVTISWTRQNGEAVKTH 270

QY 274 TKPREQVNSTYRVSVLTVLHQDMLNGEKYCKVSNALPAPIEKITSKAKGP-REPO 332
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 271 TWISSHPNATSVANGASICEDDMNSGERFTCTVTHDLPRLKQITISRPKGVALLHRP 330

QY 333 VTLTPSRDELTKQVSLTCLVKGFPSPDIAVESNGQP--ENNYKTPPVLD--SDG 387
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 331 VLLPFPARQNLRESAIIITCLVTGFSPADVFVQMMORQGPLSPKRYVTSAPMPEPPAPG 390

QY 388 SFFLSKLTVDKSRMQGNVFSQVMHEALHNHYTKSLSPGLQDETCAEQDGLD 447
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 391 RYFAHSILTVSEEMNTGETYTCVVAHEALPNRVTERTVDS-----TEGEVSHDEGEF 445

QY 448 GLMTT 452
   ||
DB 446 NLMAT 450

```

RESULT 54
S37768
Ig mu chain C region - human
C:Species: Homo sapiens (man)
C:Date: 12-Feb-1998 #sequence_revision 12-Feb-1998 #text_change 23-Jul-1999
C:Accession: S37768
R:Harindranath, N.; Donadel, G.; Sigounas, G.; Nockins, A.L.
Mol. Immunol. 30, 111-112, 1993
A:Title: Comparison of complete nucleotide sequence of the human Igm heavy chain constant
A:Reference number: S37767; MUID:93109369; PMID:8411370
A:Accession: S37768
A:Molecule type: mRNA

A:Residues: 1-453 <HAR>
A:Cross-references: EMBL:X67301; NID:938407; PIDN:CAA47714.1; PID:938408
A:Experimental source: cell line Ab 63
C:Genetics:
A:Map position: 14q32
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: glycoprotein; heterotetramer; immunoglobulin
F:21-90/Domain: immunoglobulin homology <IMM1>
F:127-199/Domain: immunoglobulin homology <IMM2>
F:237-305/Domain: immunoglobulin homology <IMM3>
F:344-415/Domain: immunoglobulin homology <IMM4>
F:14/Disulfide bonds: interchain (to light chain) #status predicted
F:28-88,134-197,244-303,351-413/Disulfide bonds: #status predicted
F:46,202,272,279,440/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:214,452/Disulfide bonds: interchain (to heavy chain) #status predicted
F:291/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 15.9%; Score 385; DB 2; Length 453;
Best Local Similarity 24.9%; Pred. No. 2.8e-16;
Matches 114; Conservative 73; Mismatches 164; Indels 106; Gaps 15;

```

QY 36 TWELTCTASQ--KKSIOFHMKNSNOIKILNQGSLFKGPSKLNDRADSRSLMDQGNFP 93
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 23 SVAVGCLAQDFLPDSITFSWKYKNSDISSTRG-----FP 57

QY 94 LIINKLKIEDSDTYCEVEDQKEVQLLVFGLTANSDFHL-----QGSLTLT 142
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 58 SVLRGKTAATSQVLPSKDV-----VMQGTDEHVCKVQHPGNKEKNVPLRVI 113

QY 150 SPSVQCR-----SPR-----GKNIOGKTLVSQ 180
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 114 SVFVPRDGFPGNPKRSKILCOATGFSPRQIQVSWLRGKQVSGVTTDQYAKESGP 173

QY 181 -----TWTCVLOQOKVEFKIDIVCPAPBPSCKTHTCP 220
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 174 TTYKVTSTLITIKESDMLSQSMFTCRV--DHRGLTFQGNASSMCPDDDTAIR----- 223

QY 221 GGPSVFLPPPKKDTLMSRPEVTCVAVDVSHEDPEKFMVYDVGVNNAKTPREO 280
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 224 -----VFALPPS-FASIFLTSTKLCLVTLDTLTYD-SVTISWTRQNGEAVKTH 277

QY 281 YNSTYRVSVLTVLHQDMLNGEKYCKVSNALPAPIEKITSKAKGP-REPOVYTLPPS 339
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 278 PNATFSAVGEASICEDDMNSGERFTCTVTHDLPRLKQITISRPKGVALLHRP 337

QY 340 RDELTKQVSLTCLVKGFPSPDIAVESNGQP--ENNYKTPPVLD--SDGSFFLYSK 394
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 338 REGNLRESAIIITCLVTGFSPADVFVQMMORQGPLSPKRYVTSAPMPEPPAPG 397

QY 395 LTVDKSRMQGNVFSQVMHEALHNHYTKSLSPG 431
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 398 LTVSEEMNTGETYTCVVAHEALPNRVTERTVDSKG 434

```

RESULT 55
S25705
Ig mu chain - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 23-Jul-1999
C:Accession: S25705
R:Patil, S.; Nau, P.
Mol. Immunol. 29, 829-836, 1992
A:Title: Isolation and sequence of a cDNA coding for the immunoglobulin mu chain of the
A:Reference number: S25705; MUID:92342148; PMID:1655860
A:Accession: S25705
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-592 <PAT>
A:Cross-references: EMBL:X59994; NID:91269; PIDN:CAA42611.1; PID:91270
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:483-554/Domain: immunoglobulin homology <IMM>

Query Match 15.9%; Score 384; DB 2; Length 592;
Best Local Similarity 27.0%; Pred. No. 4,5e-16;
Matches 123; Conservative 78; Mismatches 153; Indels 102; Gaps 23;

QY 35 DTVELTCTASQ--KSIQPHMKNSQIKILNQGSFLTKGSPKINDRADSRSLMDQNF 92
DB 161 NTVALLGCLARDVPNPSVSEFSMKFNSTVSSERFWTF-----BEVLKD-----GLMSASSQ 210
QY 93 PLIINKLIKIEDSDTY-ICEVEDQK-----EEVQL-----VFGLTANSPTHLLOG 136
DB 211 VALHSSSTFGGTDGYLCEVGQHPKGEDVGHKGVPREVEVLSPVSVFPPCNS----- 263
QY 137 QSLTITLSPSSSPSSVOCR---SPR-----GKNT-----QGSKT--- 168
DB 264 -----LSGNGNSKSLICQATDFSPKOISLWFRDGRIVSDISEGVETQSPTTYR 317
QY 169 ---LSVQLLELDGSGTMTCTVLONQKVEFIDVPCPAPRPSCDKTHRCPELLGSPS 224
DB 318 AVSLVLTIEREVLSSAVTTCOVENKKEFFQKNAS-----SSCDTTPSP--TG-- 363
QY 225 VFLPPPKPDLTLMISRTPEVTCCVVVDVSHEDPEVKFN--YVDG--VEVHNAKTKPREEQY 281
DB 364 VFTIPSPFAD-IFLRKSAKLSTGLVNTNLASYD-GLNISMHONGKALEHTY----FEHNL 417
QY 282 NSTYVSVSVLYVLHODMLNGKEYKCKVSNKALPADIETISKAG-QPREQVYTLPPSR 340
DB 418 NDTFARGAGBASCESDMESGEERYCTVAHLDLFPFEKSAISKPKDVAKPPSVYVLPTR 477
QY 341 DELT-KNQVSLTCLWKGFPSPDIAYEWESNQP--ENNYKTPPLD--SGSFFLYSKL 395
DB 478 EQLSRRESASVYCLVKGPRADVPFQWLQKGPVAKSKRYTVSSPAPEQDPSAFVHSIL 537
QY 396 TVDKSRMOQGNFSCSVNHEALHNHYTOKSLSLSPG 431
DB 538 TYTEDMSKGETTYTCVGVGHEALPHAVTERTYDKSG 573

RESULT 56
MHMM
Ig mu chain C region, membrane-bound splice form - human
C.Species: Homo sapiens (man)
C.Date: 29-Jan-1993 #sequence revision 23-Aug-1997 #text_change 22-Jun-1999
C.Accession: S16510, S09357; S16656; B26243; A02167
R.Doral, H.; Gillies, S.D.
Nucleic Acids Res. 17, 6412, 1989
A.Title: The complete nucleotide sequence of a human immunoglobulin genomic C-mu gene.
A.Reference number: S09357; MUID:8936690; PMID:2505237
A.Accession: S16510
A.Molecule type: DNA
A.Residues: 433-473 <DOR1>
A.Cross-references: EMBL:X14939
A.Note: the authors translated the codon AAC for residue 445 as Met
A.Note: the sequence of residues 1-432 was assumed to be identical with the corresponding
A.Accession: S09357
A.Molecule type: DNA
A.Residues: 1-432, 'GKPTLVNLSLWSDTAGTCY' <DOR2>
A.Cross-references: EMBL:X14940
A.Note: the authors translated the codon AAT for residue 16 as Met
A.Note: secreted splice form
R.Doral, H.
submitted to the EMBL Data Library, April 1989
A.Reference number: S16656
A.Accession: S16656
A.Molecule type: DNA
A.Residues: 1-39,'L',41-432,'GKPTLVNLSLWSDTAGTCY' <DOR3>
A.Cross-references: EMBL:X14940
R.Rabbits, T.H.; Forster, A.; Milstein, C.P.
Nucleic Acids Res. 9, 4509-4524, 1981
A.Title: Human immunoglobulin heavy chain genes: evolutionary comparisons of C-mu, C-delta
A.Reference number: A26243; MUID:82059479; PMID:6795593
A.Accession: B26243
A.Molecule type: DNA
A.Residues: 433-436,'N',438,'E',440-447,'T',449-473 <RAB>

A:Cross-references: GB:K01310; MID:q184715; PIDN:AAB59422.1; PID:q184734
C:Comment: During differentiation, B lymphocytes switch from expression of membrane-bound
S.
C:Genetics:
A:Gene: GDB:IGHM
A:Cross-references: GDB:120086; OMIM:147020
A:Map position: 14q32.33-14q32.33
A:Intron: 1/1; 105/1; 217/1; 323/1; 433/1; 477/3
C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. The IGH subunits associate into disulfide linked pentamers.
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: alternative splicing; glycoprotein; heterotrimer; immunoglobulin; membrane
F:21-90/Domains: immunoglobulin homology <IMM1>
F:127-199/Domains: immunoglobulin homology <IMM2>
F:237-305/Domains: immunoglobulin homology <IMM3>
F:344-415/Domains: immunoglobulin homology <IMM4>
F:433-473/Domains: carboxyl-terminal #status predicted <CTS>
F:14/Disulfide bonds: interchain (to light chain) #status experimental
F:128-88,134-197,244-305,351-413/Disulfide bonds: #status experimental
F:146,209,272,279/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:14/Disulfide bonds: interchain (to heavy chain) #status experimental
F:291/Disulfide bonds: interchain (to mu chain in another subunit) #status experimental

Query Match 15.8%; Score 381; DB 1; Length 473;
Best Local Similarity 23.9%; Pred. No. 5.2e-15;
Matches 116; Conservative 77; Mismatches 166; Indels 126; Gaps 17;

QY 36 TVELTCTASQ--KKSIQFMHKNQKIKILGNQSFLLKRGSKLNDRAHSRLMDQGNFP 93
DB 23 SVAVGCLADFLPDDSTTFESWKYKNQNSDISSTRG-----FP 57
QY 94 LIRKULKIEDSDTYICEVEDEKEEVOLLVFLGLTANSPDTHL-----QGSLTLYT 142
DB 58 SVLRGKVKATQGVLLPSMD-----VWGIDEHVYCKQHPRNGEKKNVPLEVI 106
QY 143 LESPRESSPSVOCR-----SPR-----GKNIGCKTLYSVQL 174
DB 107 AELPPVSVFVPPRDGFGNPRKSKLICQATGFSFPGIQVSWLRGKQVSGVTTDQVQA 166
QY 175 ELQDSS-----TWTCITLQNKQKVEFKIDYPCRAPEPSCOKT 213
DB 167 EAKESGPTTYKATSTLTILIESDWLGQSMFTCRV--DHRGTLFQGNASSMCVPDDQTAIR- 223
QY 214 HTCPHELGGPSVFLFPPEKCDTLMISRTPEVTCVVVDVSHEDPEVKKNMYDVGVENNAK 273
DB 224 -----VFATPPS-FASIFLTKSTKLTCLVLDLTYYD-SVTLSMTQNGEAVKTH 270
QY 274 TKPREQVNSTYRVVSVLTFLHODWLNKGEYKCKVSKNALPAPIEKTISKAKGP-REPQ 332
DB 271 TWISSEHPATFSAVGEASICEEDMNSGERFTCTVHTDLPSPFKQITSRPKGVALHRPD 330
QY 333 VYTLPPSRBELT-KNOVSLTCLVKGYFSPDIAVEMESNGQP--ENNYKTTPTVLD--SDG 387
DB 331 VYLLPPARQQLMRESATITCLVTGTFSPADVPQOMQGRGPLSEKVTYSPMPPEPPQPG 390
QY 388 SFFLYSKLTVDKSRMQGVNFSCSVNHEALHNHYTKSLSPQLQIDETCAEKODGELD 447
DB 391 RYFASISILTVSEENWGTGETYTC-VAHEALPNVYTERTVDKS-----TGEVSDADBEQFE 444
QY 448 GLWTT 452
DB 445 NLMAT 449

RESULT 57
538864
Ig epsilon chain C region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-May-2001
C:Accession: S38864
R:Klipp, B.; Becker, W.; Schlaak, M.
Submitted to the EMBL Data Library, November 1993
A:Description: Combination of a defined specificity and desired isotype by cloning of an

A:Reference number: S38864
 A:Accession: S38864
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-548 <R1P>
 A:Cross-references: EMBL:Z27397; NID:G416537; PIDN:CAA81788.1; PID:G940782
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 F:353-421/Domain: immunoglobulin homology <IMM>

Query Match 15.8%; Score 381; DB 2; Length 548;
 Best Local Similarity 30.6%; Pred. No. 6, 2e-16;
 Matches 103; Conservative 68; Mismatches 106; Indels 60; Gaps 16;

QY 115 KEEVLT-LVFGLTNN--SDTHLQGGSTLTLESPPSSPEVQGRSPRKNIGGGTKLTS 170
 DB 248 RSTIDLYCFYGHILINDSVSWLMDREITDTL-----AQVYLKE-EGKLASCSKLN 300
 QY 171 VSQLELDQSGTWCTVTLNQKKVEFKIDIVPCPAPEPKSCDKHTHCPPLLGGPVFLPP 230
 DB 301 ITEQOMSESPFTCKV--TSQGVVDYLAHTRRCPDHEPR-----GVITVLIIP 345
 QY 221 KPKDTLMSRPEVTCVVDV-SHDEPKFN-----WYVDGEVHNATKPRE 278
 DB 346 SPLD-LYNGAPKLTCLVVDLESEKNVVTWNOEKKTSVASQWY---TKHNN----- 394
 QY 279 EGYNSTYRVSVTLVLDHDMNGEKYCKVSNKALPAPIEKTISKAKGPREPVYTLPP 338
 DB 335 ---NATTSITSLPVAADWINGYQICLVHDPDPKIKVSIITPCQRSAPVYVPP 451
 QY 339 SRDELTKNQVSLTCLVKGPYPSDIAVEWESNGQPENN--YKTTPVLDSDG--FLYS 393
 DB 452 PEE-SEDKRITLTCLIQNFPEDISVQWLGDGKLISNSQHSSTTP-LKNSGNSRGFFIS 509
 QY 394 KLTVDKSRNQGNVSCSYMHALIN-HYTKSLSL 429
 DB 510 RLEVAKTLMTQKQPTCOVHIALQKPRLEKTISTS 546

RESULT 58
 MHMU
 Ig mu chain C region, secreted splice form - human
 C:Species: Homo sapiens (man)
 C:Date: 29-Jul-1981 #sequence revision 23-Aug-1997 #text change 22-Jun-1999
 C:Accession: S09357; S16656; S37767; A26243; A26244; I37749; I37750; A02162; B02162
 R:Dorai, H.; Gillies, S.D.
 Nucleic Acids Res. 17, 6412, 1989
 A:Title: The complete nucleotide sequence of a human immunoglobulin genomic C-mu gene.
 A:Reference number: S09357; MUID:8936690; PMID:2505237
 A:Accession: S09357
 A:Molecule type: DNA
 A:Residues: 1-452 <DOR1>
 A:Cross-references: EMBL:X14940
 A:Note: the authors translated the codon AAT for residue 16 as Met
 R:Dorai, H.
 submitted to the EMBL Data Library, April 1989
 A:Reference number: S16656
 A:Accession: S16656
 A:Molecule type: DNA
 A:Residues: 1-39, 'L', 41-452 <DOR2>
 A:Cross-references: EMBL:X14940
 R:Hardinghan, N.; Donadel, G.; Sigounas, G.; Nockins, A.L.
 Mol. Immunol. 30, 111-112, 1993
 A:Title: Comparison of complete nucleotide sequence of the human Igm heavy chain constant
 A:Reference number: S37767; MUID:93109369; PMID:8413730
 A:Accession: S37767
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-190, 'S', 192-414, 'V', 415-452 <HAR>
 A:Cross-references: EMBL:X67292; NID:G38405; PIDN:CAA47708.1; PID:G38406
 R:Rabbits, T.H.; Forrester, A.; Milstien, C.P.
 Nucleic Acids Res. 9, 4509-4524, 1981
 A:Title: Human immunoglobulin heavy chain genes: evolutionary comparisons of C-mu, C-delta
 A:Reference number: A26243; MUID:82059479; PMID:6795593

A:Accession: A26243
 A:Molecule type: DNA
 A:Residues: 1-17, 105-186, 200-259, 296-322, 339-416, 'D', 418-452 <RAB>
 A:Cross-references: GB:K01310; NID:G184715
 R:Dolby, T.W.; Devono, J.; Croce, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 77, 6027-6031, 1980
 A:Title: Cloning and partial nucleotide sequence of human immunoglobulin mu chain cDNA fr
 A:Reference number: A26244; MUID:81077306; PMID:6777778
 A:Accession: A26244
 A:Molecule type: mRNA
 A:Residues: 298-386, 436-452 <DOL>
 A:Cross-references: GB:J00257; NID:G185053; PIDN:AAA53508.1; PID:G185056; GB:J00258; NID
 R:Yanagishita, N.; Nakai, S.; Honjo, T.
 Nucleic Acids Res. 8, 5983-5991, 1980
 A:Title: Cloning of human immunoglobulin mu gene and comparison with mouse mu gene.
 A:Reference number: I37748; MUID:81124312; PMID:6450943
 A:Accession: I37749
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 347-370 <TAK1>
 A:Cross-references: EMBL:V00562; NID:G33448; PIDN:CAA33825.1; PID:G929649
 A:Accession: I37750
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 433-452 <TAK2>
 A:Cross-references: EMBL:V00563; NID:G33454; PIDN:CAA33826.1; PID:G825684
 R:Maheco, E.; Barnikol-Matanabe, S.; Barnikol, H.U.; Maheco, C.; Hilschmann, N.
 Eur. J. Biochem. 111, 275-286, 1980
 A:Title: The primary structure of the constant part of mu-chain-disease protein BOT.
 A:Reference number: A02162; MUID:81066716; PMID:6777162
 A:Contents: Mu-chain-disease protein Bot and revisions to sequence of Gal
 A:Accession: A02162
 A:Molecule type: protein
 A:Residues: 1-17, BPS', 22, 'T', 24-82, 'N', 84-90, 'Z', 92-93, 'B', 95, 'B', 97-144, 'E', 146-162, 'E'
 A:Accession: B02162
 A:Molecule type: protein
 A:Residues: 100-144, 'E', 146-162, 'E', 166-214, 'G', 216-262, 'D', 264-295, 'D', 297-414, '
 A:Note: all four combinations of the 191-Ser/Gly and 215-Val/Gly polymorphisms have been
 R:Matanabe, S.; Barnikol, H.U.; Horn, J.; Bertram, J.; Hilschmann, N.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 1505-1509, 1973
 A:Title: The primary structure of a monoclonal Igm-immunoglobulin (macroglobulin Gal), IJ
 A:Reference number: A02064; MUID:75059123; PMID:4803843
 A:Contents: annotation; Waldenström's macroglobulin Gal
 A:Note: this sequence has been revised in reference A02162
 R:Putnam, F.W.; Florent, G.; Paul, C.; Shindoda, T.; Shimizu, A.
 Science 182, 287-291, 1973
 A:Title: Complete amino acid sequence of the mu heavy chain of a human Igm immunoglobul
 A:Reference number: A02088; MUID:74005511; PMID:4742735
 A:Contents: annotation; Waldenström's macroglobulin Oa, sequence, disulfide bonds, and ce
 A:Note: this sequence differs from that shown at a number of positions; this sequence has
 C:Comment: During differentiation, B lymphocytes switch from expression of membrane-bound
 B.
 C:Genetics:
 A:Gene: GDB:IGHM
 A:Cross-references: GDB:120086; OMIM:147020
 A:Map position: 14q32.33-14q32.33
 A:Introns: 1/1; 105/1; 217/1; 323/1; 433/1
 C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa)
 hain disulfide bonds. The Igm subunits associate into disulfide linked pentamers.
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: alternative splicing; glycoprotein; heterotrimer; immunoglobulin
 F:21-90/Domain: immunoglobulin homology <IMM1>
 F:127-199/Domain: immunoglobulin homology <IMM2>
 F:237-305/Domain: immunoglobulin homology <IMM3>
 F:344-415/Domain: immunoglobulin homology <IMM4>
 F:432-452/Domain: carboxyl-terminal <CTS>
 F:14/Disulfide bonds: interchain (to light chain) #status experimental
 F:28-88, 134-197, 244-303, 351-413/Disulfide bonds: #status experimental
 F:46, 209, 272, 279, 439/Binding site: carbohydrate (asn) (covariant) #status experimental
 F:214, 451/Disulfide bonds: interchain (to heavy chain) #status experimental
 F:291/Disulfide bonds: interchain (to mu chain in another subunit) #status experimental

Query Match 15.6%; Score 376.5; DB 1; Length 452;
 Best Local Similarity 24.1%; Pred. No. 9.4e-16;
 Matches 112; Conservative 75; Mismatches 156; Indels 121; Gaps 16;

QY 36 TVELTCTASQ--KKSIOFHMKNSNQIKIGNGSFLTKGPKLNDRAISRSLMDGNFP 93
 DB 23 SAVAGCLADPLPDSITFSMKYKNNSDISSTRG-----PP 57
 QY 94 LIIKLIKIDSDTYICEVEDQKEVLLVFGLTANSDTLH-----QGSLTIT 142
 DB 58 SVLRGKVAATGQVLLPSKD-----VWQGDDEHYVCVQHPNKNKEVPLPI 106
 QY 143 LESPSPSPVOCR-----SPR-----GKNIQGGKTLVSQ 174
 DB 107 AELPKNVSVFVPRDGFPGNPKSKICQATGSPQIOGVSWLRKQVGSVTTDQV 166
 QY 175 ELQDSG-----TWCTVLQNGKVEFKIDIVCPAPEPKSCDKT 213
 DB 167 EAKESGPTTYKVTSTLTIKESDWLQGSMTCTRY--DHRGLTFQNNASMCVPDQDTAIR- 223
 QY 214 HRCPELLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAK 273
 DB 224 -----VFALPPS--FASIFLTKSTKLTCLVTLDTTYD--SVTISWTRQGEAVKTH 270
 QY 274 TKPREQVNSTYRVSVLTVLHODMNGKCYKCKVSNKALPAPIEKTISKAKGQP--REQ 332
 DB 271 TINSSEHPNATPSAVEASICEEDMNSGERFTCTVHTDLPFLKTTISRPKVALHRD 330
 QY 333 VYTLPSRDELTKNGVSLTCLVKGFYPSDIAVEWESNGQ--ENNYKTTTPVLD--SDG 387
 DB 331 VYTLPPARQQLNRESATITCLVGTGSPADVFQVMQROQPLSPKCYVTSAPWPEQVAG 390
 QY 388 SFPLYSKLTIVYDSRMQGVNFSGVNHEALAHNYTQKSLSLSPG 431
 DB 391 RYFAHSILTVSEEWNTGETTYTC--VAHEALPNRVTERTVYDKSTG 433

RESULT 59
 S03186
 Ig heavy chain C region - African clawed frog (fragment)
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Sep-1999
 C:Accession: S03186
 R:Haite, R.N.; Shambloct, M.J.; Amemiya, C.T.; Litman, G.W.
 Nucleic Acids Res. 17, 1776, 1989
 A:Title: A second Xenopus immunoglobulin heavy chain constant region isotype gene.
 A:Reference number: S03186
 A:Accession: S03186
 A:Molecule type: mRNA
 A:Residues: 1-448 <HA1>
 A:Cross-references: EMBL:X13779; NID:g64827; PID:g64828
 A:Note: the authors translated the codon TTT for residue 9 as Ser and ATG for residue 34
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keyword: heterotetramer; immunoglobulin

Query Match 15.5%; Score 374; DB 2; Length 448;
 Best Local Similarity 26.8%; Pred. No. 1.3e-15;
 Matches 117; Conservative 57; Mismatches 144; Indels 118; Gaps 16;

QY 80 ADSRRSLMDQGNFPLIIKLIK--IEDSD-----TYICEV---EDQ 114
 DB 35 ADPLDQWNGDSTTTCIKMRPVLSDVLDYLTSSQLTLASEMKNSTYKCVVINYNT 94
 QY 115 KEKVLLVFGLTANSDTLHQQGSLTTLSPSPGSP----- 151
 DB 95 KQEKSLKVLPCMA-----PHYQLFQSPCMSDAISRAQHININATLDCIINFY 145
 QY 152 -----SVQCRSPRGKNIQGGKTLVSQLELQDSGW-----TCVVLON 189
 DB 146 HQGIKVKMLVNGQDVSAASVPTPKRTDGYVSSSQRLIL-KGMWNGTQVYSCIVTHT 204
 QY 190 QKKVEFKIDIVCPAPEPKSCDKTHRCPELLGSPSVFLFPKPKDTLMISRTPEVTCV- 248

DB 205 SNTTTIANISQC-----TEQCHDNL---QVPLPTFPHD-LYFSRNKATYCLVS 250
 QY 249 -----VDVSHEDPEV-KFNMYVDGVEVNAKTPREQVNSTYRVSVLTVLHODMNL 300
 DB 251 SMKTIENPDISWERKAGNLEFVTEDPVLAHD-----NGTYSVASIILSVCAEDWS 300
 QY 301 GKEYCKVSNKALPAPIEKTISKA-KGQPREQVYTLPPSRDELTKNQ-VSLTCLVKGPY 358
 DB 301 GDEKFCYRSQPLSPSPVKTTIFKQNEGTGPKADVLLPSPSAQELLQGEWVTLTCTVTGN 360
 QY 359 PSDIAVEWESNG---QPENNYKTTTPPVLSDSGSFPLYSKLTVDKSRMOQGVNFSGVNHE 415
 DB 361 PEKFIQWNGGVSISEDKFINTVPMKSDGEGTYFIYSKLAIPAKMNGQDVFTCVGHE 420
 QY 416 ALAHNYTQKSLSLSPG 431
 DB 421 ALPLYITQDSIDKSSG 436

RESULT 60
 M8KSM
 Ig mu chain C region, membrane-bound form - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 31-Oct-1980 #sequence_revision 30-Jun-1991 #text_change 22-Jun-1999
 C:Accession: A02167; A37517; B02166
 R:Barly, P.; Rogers, J.; Davis, M.; Calame, K.; Bond, M.; Wall, R.; Hood, L.
 Cell 20, 313-319, 1980
 A:Title: Two mRNAs can be produced from a single immunoglobulin mu gene by alternative R
 A:Reference number: A02167; MUID:80222874; PMID:6771020
 A:Accession: A02167
 A:Molecule type: DNA
 A:Residues: 436-476 <EAR>
 A:Cross-references: GB:V00816; GB:J00444; NID:g52343; PID:CAA24197.1; PID:g52344
 R:Rogers, J.; Barly, P.; Carter, C.; Calame, K.; Bond, M.; Hood, L.; Wall, R.
 Cell 20, 303-312, 1980
 A:Title: Two mRNAs with different 3' ends encode membrane-bound and secreted forms of im
 A:Reference number: A37517; MUID:80222873; PMID:6771019
 A:Contents: MOPC 104E
 A:Accession: A37517
 A:Molecule type: mRNA
 A:Residues: 410-476 <ROG>
 A:Cross-references: GB:V00821; NID:g52355; PIDN:CAA24202.1; PID:g817972
 R:Kawakami, T.; Takahashi, N.; Honjo, T.
 Nucleic Acids Res. 8, 3933-3945, 1980
 A:Title: Complete nucleotide sequence of mouse immunoglobulin mu gene and comparison wit
 A:Reference number: A02166; MUID:81076590; PMID:6255422
 A:Accession: B02166
 A:Molecule type: DNA
 A:Residues: 1-435, 'GKPTLVNVSILMSDTGTCV' <KAW>
 C:Comment: The sequence of residues 1-409 was assumed to be identical with the correspon
 C:Comment: During differentiation, B lymphocytes switch from expression of membrane-bound
 B.
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobu
 F:21-91/Domain: immunoglobulin homology <IMW1>
 F:129-201/Domain: immunoglobulin homology <IMW2>
 F:239-307/Domain: immunoglobulin homology <IMW3>
 F:346-417/Domain: immunoglobulin homology <IMW4>
 F:435-476/Domain: carboxyl-terminal <CNS>
 F:14/Disulfide bonds: interchain (to light chain) #status predicted
 F:28-89,136-199,246-305,353-415/Disulfide bonds: #status predicted
 F:46,211,243,258,281/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:216/Disulfide bonds: interchain (to heavy chain) #status predicted
 F:293/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 15.5%; Score 373.5; DB 1; Length 476;
 Best Local Similarity 25.2%; Pred. No. 1.5e-15;
 Matches 121; Conservative 87; Mismatches 157; Indels 115; Gaps 23;

QY 37 VELTCTASQ--KKSIOFHMKNSNQIKIGNGSFLTKGPKLNDRAISRSLMDGNF-- 92

```

Db      24 VAMGCLADFLPSTISFTMYNQNTTEVIQIGIRTEPT-----LRTGKXLA 68
Qy      93 ---PILINKLKIENSDPY-ICEVEDQKEVQLVYGLTANSPTL-----LOGSLTLTLE 144
Db      69 TSQVLSPEKSLLEGSDYLCKIH-----YG-GKNLDLHVLPVAVENKPNVNF 117
Qy      145 SPP-----GSSP---SVOCR-----SPR-----GKNIQGG----- 166
Db      118 VPPDGFSGPAPRPSKSLCEATNTPPKRITVSWLKDGLVSSGFTTDPVTIENKSTPQT 177
Qy      167 ---KTLVSQLELQDSTGTCTVQLQNKQVPEFKIDIVCPAPEKSCDKHTTCELLIG 222
Db      178 YKVIITLSTLSEIDMLNLNLVYTCRV--DHRGLTFLKLVNVSSTCAASPST-----DIL- 225
Qy      223 PSVFLFPKPKDPTLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEVNAKTKPREBYN 282
Db      226 --TFIPSPAD-IFLSANLTCVLSNLATYE-TLNISMSQSGEPLETIKIMESHPN 281
Qy      283 STYRVSVLTVLHODWLNKGEYKCKVSNKALPAPIEKTISKAKQPRE-----POVYTLF 337
Db      282 GTFSAKGVASVCEVEDMNRKFEVCTVTHRDLPSPQKRISK-----PNEVHGHPRAVILLP 337
Qy      338 PSRDELTKNOVSLTCLVKGFPYSDIAVESNGQ--PENNYKTPPVLD--SPGSEFLY 392
Db      338 PAREQLMREBAYTCTLVKGFPSPADISVQWLQRGQLLPQEKVYTSAPPEPGAGGFYETH 397
Qy      393 SKLTVDKSRMOQGVNFSCVWHEALHNHYTOKSLSLSPGLDDETCAPADQDELGLMTT 452
Db      398 SILVTVEEMNSGETTYTCVGHGHALPLVTERTVDKS-----TEGEVNAEBEGENLWTT 452

```

RESULT 61

```

EHRT
Ig epsilon chain C region - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 17-Dec-1992 #sequence_revision 17-Dec-1992 #text_change 16-Jul-1999
C/Accession: A93442; A90937; A02143
R/Hellman, L.; Petersen, U.; Engstrom, A.; Karlsson, T.; Bennich, H.
Nucleic Acids Res. 10, 6041-6049, 1982
A/Title: Structure and evolution of the heavy chain from rat immunoglobulin E.
A/Reference number: A93442; MUID:83064537; PMID:6292865
A/Accession: A93442
A/Molecule type: mRNA
A/Residues: 1-429 <HEL>
A:Experimental source: strain LOU/c/Mel, immunocytoma IR2
R:Kindersvogel, W.R.; Reddy, E.P.; Moore, J.M.; Faust Jr., C.H.
DNA 1, 335-343, 1982
A/Title: A cloned cDNA probe for rat immunoglobulin epsilon heavy chain: construction,
A/Reference number: A90937; MUID:83182019; PMID:6820340
A/Contents: myeloma IR162
A/Accession: A90937
A/Molecule type: mRNA
A/Residues: N, 169-307, U, 309-342 <KIN>
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap)
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C/Superfamily: immunoglobulin C region, immunoglobulin homology
C/Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F:118-166/Domain: immunoglobulin homology <IM1>
F:118-166/Domain: immunoglobulin homology <IM2>
F:223-291/Domain: immunoglobulin homology <IM3>
F:321-398/Domain: immunoglobulin homology <IM4>
F:46,99,170,240,265,369,419/Binding site: carbohydrate (Aen) (covalent) #status predicted

```

```

Query Match      15.3%; Score 370.5; DB 1; Length 429;
Best Local Similarity 31.5%; Pred. No. 2e-15;
Matches 88; Conservative 62; Mismatches 102; Indels 27; Gaps 10;
Qy      160 GKNIQGGKTLVSQLELQDSTGTCTVQLQNKQVPEFKIDIVCPAPEKSCDKHTTCELL 219
Db      160 GKLAISTYRLNITQOGMSESTFTCKVTSQGE--NYMAHTRRCSDDER----- 206
Qy      220 LGBSVFLFPKPKDPTLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEVNAKTKPREE 279

```

```

Db      207 --GVITVLIPEPSPLD--LYENGTEPKLTCLVLDESE--NITVWVERKKSIGSASQSTX 262
Qy      280 QNSTYRVSVLTVLHODWLNKGEYKCKVSNKALPAPIEKTISKAKQPREPOVYT-LPP 338
Db      263 HHNAITSTISLTPDADWTEGEGYQCRVDHPHPKDIVSITVAPKRSAPVEYVFLPP 322
Qy      339 SRDELTKNOVSLTCLVKGFPYSDIAVEW--ESNQPENNYKTPPVLDSDG---PFLYS 393
Db      323 EEEF--KDKRTITCLIONFFPEDISVQWLQDCLKLIPSSQSTTTP-LKNYNSNRFPIFS 379
Qy      394 KLTVDKSRMOQGVNFSCVWHEALHN-HYTQKSLSLSPG 431
Db      380 RLEVTKALMTQTKQFCTCRVTHEALREPRKLERISKSLG 418

```

RESULT 62

```

MHREM
Ig mu chain C region, membrane-bound form - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 04-Dec-1986 #sequence_revision 30-Jun-1991 #text_change 23-Aug-1997
C/Accession: A02165; A02164
R/Bernstein, K.E.; Alexander, C.B.; Reddy, E.P.; Mage, R.G.
J. Immunol. 132, 490-495, 1984
A/Title: Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain of V-Ha2 aJ
A/Reference number: A02164; MUID:84088930; PMID:6418803
A/Contents: a2 allotype
A/Accession: A02165
A/Molecule type: mRNA
A/Residues: 439-479 <BE2>
A/Accession: A02164
A/Molecule type: mRNA
A/Residues: 1-438, 'GKPTLVNVSILMSDASTCY', <BER>
A/Note: the sequence of residues 1-438 was assumed to be identical with the corresponding
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap)
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin
F:21-92/Domain: immunoglobulin homology <IM1>
F:130-202/Domain: immunoglobulin homology <IM2>
F:242-320/Domain: immunoglobulin homology <IM3>
F:349-420/Domain: immunoglobulin homology <IM4>
F:439-479/Domain: immunoglobulin homology <IM5>
F:14/Disulfide bonds: interchain (to light chain) #status predicted
F:28-90,137-200,249-308,356-418/Disulfide bonds: #status predicted
F:46,114,212,261,277,284/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:219/Disulfide bonds: interchain (to heavy chain) #status predicted
F:236/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

```

```

Query Match      15.3%; Score 369.5; DB 1; Length 479;
Best Local Similarity 23.8%; Pred. No. 2.7e-15;
Matches 123; Conservative 79; Mismatches 153; Indels 161; Gaps 23;
Qy      21 LAIQGNKVTYLGKGDVETLTCTASQ--KKSIOGHM--KNSNQIKILNQGSFLTKGPSKL 76
Db      17 ALTDGNLVAMG-----CLARDFLPSSVTFSWSPKNNSEI----- 50
Qy      77 NDRADRSRLMDGNFPLIKNK-----IDSPTY-ICEVEDQKEVQL 121
Db      51 ----SSRTV---RTFVVRKGDKNMATSGVLYPSKVDLTGETEYLVCKVQHSNSRDLR 102
Qy      122 VFGITANSPTLHLOGOSLTTLTSPSPSSPSVOCR-----SPR-- 159
Db      103 V-----SFVDSLPFPAVSVFIPPRDSFSSGTRKSLICQATFSFSL 147
Qy      160 -----GKNIQGG-----KTLVSQLELQDSTGW-----TCVY----L 187
Db      148 SVSWLRDQGVESGVLTLPVEAEITKGAGPATFSSISWLTITTESDMLSQSLYTCDVDRGI 207
Qy      188 QNKQVPEFKIDIVCPAPEKSCDKHTTCELLGSPVFLFPKPKDPTLMISRTPEVTCV 247
Db      208 FPDKNVMSSECSCTTSP-----GIQVFPFAPSADT-FLSKARLCL 250
Qy      248 VVDVSHEDPEVKFNWYVDGVEVNAKT-----KPREQYNSTYRVSVLTVLHODWLNKG 302

```

```
Db      251 VTDLTTYG--SLNLSW-----ASHNGKALDTHNMITESHNPATFSANGASVCAEDMESE 304
Qy      303 EYKCKVSKNALPAPLEKTIKSKAKGQPR-POVYTLPPSDELT-KNOVSLTCLVKGFPYS 360
Db      305 QETCTVTHADLPFLLHHTTISKSREVAKHPAVYVLPAREQLVLRSAATVTCLVGFGFSPA 364
Qy      361 DIAVESNGQPR--ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMQGQNVFSCVMEHA 416
Db      365 DVFVQMQGQPRUSDKTYTSAPAPRQAPGLYFTHTSTLTVEEDMNSGETTFCVVGHEA 424
Qy      417 LHNHYTKSLSLSPGLQDDETCANQDGLDGLMTT 452
Db      425 LPHMYTERTVDS-----TEGEVGAEEEGFENLMTT 455

RESULT 63
MHGQ
Ig mu chain C region - dog (tentative sequence)
C:Species: Canis lupus familiaris (dog)
C>Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C:Accession: A93131; A94246; A02169
R:McCumber, L.U.; Capra, J.D.
Mol. Immunol. 16, 565-570, 1979
A:Title: The complete amino-acid sequence of a canine mu chain.
A:Reference number: A93131; MUID:80077682; PMID:117259
A:Contents: myeloma protein MOO
A:Accession: A93131
A:Molecule type: protein
A:Residues: 1-177 <MCC>
R:Wasserman, R.L.; Capra, J.D.
Science 200, 1159-1161, 1978
A:Title: Amino acid sequence of the Fc region of a canine immunoglobulin M: interspecies
A:Reference number: A94246; MUID:78180587; PMID:653360
A:Contents: MOO
A:Accession: A94246
A:Molecule type: protein
A:Residues: 178-450 <WAS>
C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (lambda)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin
F:20-89/Domain: immunoglobulin homology <IMM1>
F:126-196/Domain: immunoglobulin homology <IMM2>
F:234-302/Domain: immunoglobulin homology <IMM3>
F:341-412/Domain: immunoglobulin homology <IMM4>
F:430-450/Domain: Carboxyl-terminal <CTS>
F:14/Disulfide bonds: Interchain (to light chain) #status predicted
F:27-87,133-194,241-300,348-410/Disulfide bonds: #status predicted
F:50,206,269,276,437/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:211,449/Disulfide bonds: Interchain (to heavy chain) #status predicted
F:288/Disulfide bonds: Interchain (to mu chain in another subunit) #status predicted

Query Match      15.3%; Score 369; DB 1; Length 450;
Best Local Similarity 24.1%; Pred. No. 2,7e-15;
Matches 119; Conservative 80; Mismatches 139; Indels 156; Gaps 20;

Qy      36 TWELTCTASQ--KKSIOFWKNSNOIKLGNQGSFLTKGPSKLANDRADSRSLMDQGNF 93
Db      22 TYAMGCLADFLPGSITFSMKYEBLSAINSTRG-----FP 56
Qy      94 LIIKLNK-----TEDSTYI-CEVE---DQKEVOLVPGLTANSDTHLL 134
Db      57 SYLRGKQVATQVQLPSVDITIGTDEHIVCKRHSBGBKQKVBPLPV----- 105
Qy      135 QGQSLTLTLESPGSSPSVQCR-----SPR-----GKNIQGGKT 168
Db      106 -----LTL--PREVSGFIPPRDAFRGPRKRSQQLICQAGSFSPROWSLARDKQIESGT 157
Qy      169 LSVSQLELQDSG-----TWTC-----TVLNQKQVVEFKIDIV 200
Db      158 TWEVZAKZAKSGPTTYKVTSMLTIGEDAWLSQSVFCKVEHRLGLTFQNASSM----- 210
```

```
Qy      201 PCPAPRPSCDKTHTCPELLGSPVFLPFPKQOTIMISREPEVTCVVVDVSHEDPEVKF 260
Db      211 -CTSDQPV-----GISIFITPPS-FAFINTKAKSLCLVTDLATVY-SVTI 254
Qy      261 NMVYDGVENHNAKTPREBOYNSTYRVSVLTVLHQMANGKEYCKVSKNALPAPIEKT 320
Db      255 SWTRENGALKTHTNISSEHPNGRTFSANGEAIVCEWESGQGFCTVHTHDLPSVLKQT 314
Qy      321 ISKAGQ-OPREPOVYTLPPSDEL-TKQVSLTCLVKGFPYSDIAVESNGQ--PENNY 376
Db      315 ISRPKVAVVHMSVYVLPSPSREQLDRESATLSCLVTGYSPPDVVQVQKQVPVPSY 374
Qy      377 KTPPVLD--SDGSFPLYSKLTVDKSRMQGQNVFSCVMEHALHNHYTKSLSLSPG--- 431
Db      375 VTSAPMPRQAPGLYFAHSILTVSEEMNAGETTYCVVAHESLPKRYTERSVDSGTGPT 434
Qy      432 -----LQDETCAE 440
Db      435 LYNVSLVSDPAGZ 448

RESULT 64
MHRHUR
Ig mu heavy chain disease protein (Bot) - human
C:Species: Homo sapiens (man)
C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jul-1999
C:Accession: A02163
R:Barnikol-Watanabe, S.; Mihaesco, E.; Mihaesco, C.; Barnikol, H.U.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 365, 105-118, 1984
A:Title: The primary structure of mu-chain-disease protein BOT. Peculiar amino-acid sequ
A:Reference number: A02163; MUID:84184186; PMID:6425189
A:Accession: A02163
A:Molecule type: protein
A:Residues: 1-391 <BAR>
C:Comment: This protein has no V region homology or CH1 region.
C:Genetics:
A:Gene: GDB:IGHM
A:Cross-References: GDB:120086; OMIM:147020
A:Map position: 14q32.33-14q32.33
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin; transmembrane pro
F:1-41/Domain: pre-C <VAR>
F:43-391/Domain: Ig mu chain C region, secreted form <IGM>
F:65-137/Domain: immunoglobulin homology <IMM1>
F:175-243/Domain: immunoglobulin homology <IMM2>
F:282-353/Domain: immunoglobulin homology <IMM3>
F:147,210,217,378/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      15.1%; Score 365.5; DB 1; Length 391;
Best Local Similarity 28.0%; Pred. No. 3.7e-15;
Matches 105; Conservative 65; Mismatches 132; Indels 73; Gaps 15;

Qy      110 EYED-----QKEVOLVPGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCR----- 156
Db      18 EAEEDRIIEBEARL-----SGRD---MQVTSQPIALPDKVSYFVPRPDGFGENPRKS 68
Qy      157 -----SPR-----GKNIQGGKTVSVSGLQDSG-----TWTCVLIQ--- 189
Db      69 KLICQATGFSRQIENVSWLRGKGQVSGVTVDVEAEAKESGPTTYKYTSTLTITIESDWL 128
Qy      190 -QKVFYKIDIVPCPAPRPSKDKTHTCPELLGGS-----VFLFPKPDOTLMISRTP 242
Db      129 GQSMFTGVDRHGLTFQNAS-----MCGPDQTAIRVFPIPS-FAFISLTST 178
Qy      243 EVTCVVVDVSHEDPEVKENMYVDGVEVNAKTPREBOYNSTYRVSVLTVLHQMANGK 302
Db      179 KLTCLVTDLTYYD-SVTISWRQGEAVKTHTNISEHPNATFSAVGSASICEDDWDGGE 237
Qy      303 EYKCKVSKNALPAPLEKTIKSKAKGQPR-REPOVYTLPPSDELT-KNOVSLTCLVKGFPYS 360
Db      238 RFTCTVHTDLPSPKQTSRPRKQVLAHRPPVYLVPAREQLVLRSAATITCLVVGFGFSPA 297
Qy      361 DIAVESNGQPR--ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMQGQNVFSCVMEHA 416
```


A>Title: Nucleotide sequence of a cloned cDNA corresponding to secreted mu chain of mouse
A.Reference number: A26240; MUID:81165562; PMID:6260591
A.Contents: TEPC183
A.Accession: A26240
A.Molecule type: mRNA
A.Residues: 1-225, 'N', 227-257, 'S', 259-367, 'K', 369-455 <AUF>
R.Kehry, M.; Sibley, C.; Fuhrman, J.; Schilling, J.; Hood, L.E.
Proc. Natl. Acad. Sci. U.S.A. 76, 2932-2936, 1979
A.Reference number: A26241; MUID:79223904; PMID:111247
A.Contents: annotation; MOPC 104E
A.Note: This sequence has been revised in reference A02039. Carbohydrate binding sites a
Biochemistry 21, 5415-5424, 1982
A>Title: Complete amino acid sequence of a mouse mu chain: homology among heavy chain co
A.Reference number: A02039; MUID:83075344; PMID:6816276
A.Contents: MOPC 104E
A.Accession: B02039
A.Molecule type: protein
A.Residues: 1-77, 'N', 79-100, 'Q', 102-225, 'N', 227-257, 'T', 259-367, 'K', 369-455 <KEH>
C.Genetics:
A.Introns: 1/1; 106/1; 219/1; 325/1
C.Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka)
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 16
C.Superfamily: immunoglobulin C region; immunoglobulin homology
C.Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobu
F:21-91/Domain: immunoglobulin homology <IMM1>
F:129-201/Domain: immunoglobulin homology <IMM2>
F:239-307/Domain: immunoglobulin homology <IMM3>
F:346-417/Domain: immunoglobulin homology <IMM4>
F:436-455/Domain: Carboxyl-terminal <CTG>
F:14/Dsulfide bonds: Interchain (to light chain) #status predicted
F:28-69/Dsulfide bonds: Interchain (to heavy chain) #status predicted
F:46 211 243 281 442/Binding site: carbohydrate (asn) (covalent) #status experimental
F:136-199 246-305 353-415/Dsulfide bonds: #status predicted
F:216 454/Dsulfide bonds: Interchain (to heavy chain) #status predicted
F:293/Dsulfide bonds: Interchain (to mu chain in another subunit) #status predicted

Query Match 15.0%; Score 362; DB 1; Length 455;
Best Local Similarity 25.3%; Pred. No. 7.2e-15;
Matches 116; Conservative 84; Mismatches 149; Indels 110; Gaps 22;

37 VELTCTASQ--KKSIOFHKNNSQIKILNOSFLTGKPSKLNDRADSRSLDQGNF-- 92
24 VAMGCLARFLPSTISFTNNYQNTNEVIGIRTFP-----LRTGSKYLA 68
93 ---PLIINKLKIEDSPTY-ICEVEDQKEVQLVFGLTANSSTHL---LQCSLTILE 144
69 TSQVLSPKSILGSDSEVLVCKIH-----Y6-GKGRDLHVPAPVAEMNPVNVF 117
145 SPP-----GSSP-----SVQR-----GKNIQG----- 166
118 VPPROGSGPAPKSKLICEATNFTPKPTVSMKDKVSGFPTDPVTIENKSTPQT 177
167 ----KTLVSQLELODPSGTWCTVLONOKKVEFKIDIVCPAPKSCDKHTCPELGG 222
178 YKVISLTILSEIDMLNLYTKRV--DHGGLFLKVVSTCAASPT-----DIL-- 225
223 P6VFLPPPKDQTLMI8TRPEVTCVVDVSHDEPKVFNWYDGEVHNAAKTRPREOYN 282
226 --TFITPSPFAD-IFLSKSNLTLCLVSNLATYE-TLINISMA5GSEPLETKIKIMESH 281
283 STYRVSVTLVLAHQDLNKEKCKRKNALPAPLEKITSKAKGQPRE-----POVYTLR 337
282 GFPSAKGAVSVCEDWNNRKEFCVTHDLPSPQCKFTSK---PNEVHKHPPAYLLP 337
338 PARDELT-KNOVSLTCLVKGFPSPDIAVEMESNGQ--PENNAKTPPVLD--SDGSFLY 392
338 PARQOLNRESATVTLVGVGSPADISVGMLOKGLLPBPKVYTAHPMBEPAPGFYTH 397
393 SKLTVDKSRMOQGNVFC5VMEBALHNHYTOKSLISLSPG 431
398 SILTVTEEMNSGETYTCVVGHEALPHLTVERTVDKSTG 436

RESULT 68
136948
Ig epsilon-chain - chimpanzee (fragment)
C.Species: Pan troglodytes (chimpanzee)
C.Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 21-Jan-2000
C.Accession: 136948
R.Sakoyama, Y.; Hong, K.
Proc. Natl. Acad. Sci. U.S.A. 84, 1080-1084, 1987
A>Title: Nucleotide sequences of immunoglobulin epsilon genes of chimpanzee and oranguta
A.Reference number: 136948; MUID:87147196; PMID:3103123
A.Accession: 136948
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-426 <RES>
A.Cross-References: GB:M15398; NID:9176797; PIDN:AAA35416.1; PID:9176798
C.Genetics:
A.Introns: 103/1; 209/1; 317/1
C.Superfamily: immunoglobulin C region; immunoglobulin homology
F:336-405/Domain: immunoglobulin homology <IMM>

Query Match 15.0%; Score 361.5; DB 2; Length 426;
Best Local Similarity 28.9%; Pred. No. 7.2e-15;
Matches 102; Conservative 54; Mismatches 136; Indels 61; Gaps 10;

90 GNFPLIINKLKIEDSPTYICEVEDQKEVQLVFGLTAN--SDTHLQCSLTILTLESPP 147
123 GHFPPTQLL-----CLVSGYTGTTNITWLEDGQVWDVLL----- 158
148 GSSPSPVQCRSPRGKNIQGGKTLVSQLELODPSGTWCTVLONOKKVEFKIDIVCPAPBP 207
159 ----STASATQEGELASQSELTLSQKWMLDRTYTCQVYQGGFFE----- 201
208 KSCDKHTHCPBL-LGSPVFLPPPKDQTLMI8TRPEVTCVVDVSHDEPKVFNWYD 266
202 ---DSTKCCADSNPRGSAVLSRPSFPD-LFIRKSPITTCVVDLAPSKGTNLTWSPRAS 257
267 VEVNNAKTRPREOYNSTYRVSVTLVLAHQDLNKEKCKRKNALPAPLEKITSKAKG 326
258 GPNVHSHRKKQKQKNGTLVTSTLPVGTWRMBEYVQCKVTHHLPRAVLRSTKSG 317
327 QPREPOVYTL-----P6RDELTKNQVSLTCLVKGFPSPDIAVEMESN--GQPENNYKTT 379
318 PRAAEVYAFATPREBGRDKRT-----LACLIQNFMEDISVGMLEHVEQLPDRHSTT 372
380 P6VLDSGSPFLYKLTVDKSRMOQGNVFC5VMEBALHNHYT-OXSLISLSPG 431
373 QPHKTKGSGFVFSRLVETRAEMEQKDEFICRAVHEAASPSQTVQRTVSNVG 425

RESULT 69
EHHU
Ig epsilon chain C region - human
C.Species: Homo sapiens (man)
C.Date: 31-Mar-1981 #sequence_revision 13-Jun-1983 #text_change 16-Jul-1999
C.Accession: A22771; A23195; PH1214; A93431; A90824; A94418; B93933; S02438; A53116; C46
R.Flanagan, J.G.; Rabbits, T.H.
EMBO J. 1, 655-660, 1982
A>Title: The sequence of a human immunoglobulin epsilon chain constant region gene
A.Reference number: A22771; MUID:84236029; PMID:6234164
A.Accession: A22771
A.Molecule type: DNA
A.Residues: 1-428 <FLA>
A.Cross-References: GB:L00022; GB:V00555; NID:9185035
R.Ueda, S.; Nakai, S.; Nishida, Y.; Hiseajima, H.; Honjo, T.
EMBO J. 1, 1539-1544, 1982
A>Title: Long terminal repeat-like elements flank a human immunoglobulin epsilon pseudog
A.Reference number: A23195; MUID:84207910; PMID:6327276
A.Accession: A23195
A.Molecule type: DNA
A.Residues: 2-428 <END>
A.Cross-References: GB:J00222; NID:9184755
R.Zhang, K.; Saxon, A.; Max, E.E.

J. Exp. Med. 176: 233-243, 1992

A>Title: Two unusual forms of human immunoglobulin E encoded by alternative RNA splicing

A:Reference number: PH1214; MUID:92308839; PMID:1613458

A:Accession: PH1214

A:Molecule type: DNA

A:Residues: 320-428 <ZHA>

A:Cross-references: EMBL:X63693; GB:S38668; NID:g32987

R:Seno, M.; Kurokawa, T.; Ono, Y.; Onda, H.; Saeeda, R.; Igarashi, K.; Kikuchi, M.; Sugita, N.

A>Title: Molecular cloning and nucleotide sequencing of human immunoglobulin epsilon chain

A:Reference number: A93491; MUID:83168897; PMID:6300763

A:Accession: A93491

A:Molecule type: mRNA

A:Residues: 1-428 <SEN>

A:Cross-references: GB:L00022; GB:J00227; GB:V00555; NID:g185035

R:Max, E.E.; Batley, J.; Ney, R.; Kirsch, I.R.; Leder, P.

A>Title: Duplication and deletion in the human immunoglobulin epsilon genes.

A:Reference number: A90824; MUID:83001945; PMID:6288268

A:Accession: A90824

A:Molecule type: DNA

A:Residues: 1-358, 'L', 360-428 <MAX>

A:Cross-references: GB:J00222; NID:g184755

A>Note: this sequence difference may be due to polymorphism

R:Bennich, H.H.; Johanson, S.G.O.; Von Bahr-Lindstrom, H.

A>Title: Cloning and sequence determination of the gene for the human immunoglobulin epsilon chain

A:Reference number: A93933; MUID:83065234; PMID:6815656

A:Accession: A93933

A:Molecule type: mRNA

A:Residues: 1-40; 68-114; 427-428 <KEN>

A:Cross-references: GB:L00022; NID:g185035

R:Ikeyama, S.

A>Title: Purification and characterization of a recombinant human IgE Fc-epsilon fragment

A:Reference number: S02438; MUID:88083554; PMID:3121387

A:Accession: S02438

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 98-352 <IKE>

R:Zhang, K.; Max, E.E.; Cheah, H.K.; Saxon, A.

J. Biol. Chem. 269: 456-462, 1994

A>Title: Complex alternative RNA splicing of epsilon-immunoglobulin transcripts produces

A:Reference number: A53116; MUID:94103254; PMID:8276835

A:Accession: A53116

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 320-428 <ZH2>

A:Experimental source: myeloma U266-derived cell line AF-10

A>Note: sequence extracted from NCBI Backbone (NCBIN:141701, NCBI:P:141702)

R:Hellman, L.

Eur. J. Immunol. 23: 159-167, 1993

A>Title: Characterization of four novel epsilon chain mRNA and a comparative analysis of

A:Reference number: A46536; MUID:93122085; PMID:8419166

A:Accession: A46536

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 382-426 <HEL>

A:Cross-references: GB:S55273; NID:g263166; PIDN:AA24857.1; PID:g263167

A:Experimental source: B cell myeloma U-266

A>Note: sequence extracted from NCBI Backbone (NCBI:P:125297)

A:Accession: A46536

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 382-391 <HE2>

A:Cross-references: GB:S55276; NID:g263168; PIDN:AA24858.1; PID:g263169

A:Experimental source: B cell myeloma U-266
A>Note: sequence extracted from NCBI backbone (NCBIP:125299)
A:Accession: A46536
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 401-428 <HE3>
A:Cross-references: GB:S53497; NID:g263162; PIDN:AA824855.1; PID:g263163
A:Experimental source: B cell myeloma U-266
A>Note: sequence extracted from NCBI backbone (NCBIP:123483)
C:Genetics:
A:Gene: GDB:IGHB
A:Cross-references: GDB:119335; OMIM:147180
A:Map position: 14q32.33-14q32.33
A:Insertion: 1/1; 104/1; 211/1; 339/1
C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobulin
F:128-87/Domain: immunoglobulin homology <IM1>
F:128-195/Domain: immunoglobulin homology <IM2>
F:132-301/Domain: immunoglobulin homology <IM3>
F:138-407/Domain: immunoglobulin homology <IM4>
F:154/Disulfide bonds: interchain (to light chain) #status predicted
F:115-105;29-85;135-193;239-289;345-405/Disulfide bonds: #status predicted
F:211;49;99;146;252;275/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:121;209/Disulfide bonds: interchain (to heavy chain) #status predicted

Query Match 15.0%; Score 361.5; DB 1; Length 428;
Best Local Similarity 28.9%; Pred. No. 7,2e-15;
Matches 102; Conservative 53; Mismatches 137; Indels 61; Gaps 10;

QY 90 GNEPLILNKLIEDSDTYICEVEDQKEVQLVFLGLTAN--SDTHLQGSGLTLESPP 147
Db 125 GHFFPTIIDL-----CLVSGYTPGTINITWLEDGQVMDVL----- 160
QY 148 GSSPEVQGRSPRKNIQGGKTLVSQLELDSDGTWCTVLQNGKVEFKIDYPCAPAP 207
Db 161 ---STASTQGBELASTQSELTLQKHWLSRPTYCQVLYOGHTE----- 203
QY 208 KSCDHTHCPEL-LGSPSVFLFPKPKDTLMSRTPREVCVVVDVSHEDPEVFNMYVDG 266
Db 204 ---DSTKCCADBNPRGVAAYLSRSPFD-LFRKSTPTICLVVDLAPSKTVMLTWGRAS 259
QY 267 VEHNNAKTPREBYQNVSTYRVVSVLTVLHQMNLNGEKYCKVSNKALPADIEKTIISKAG 326
Db 260 GKPVNHSSTRKEKQNGTLTVSTLPVGTWDVIEGTYQCRVTHPHLRLALMSTRTSG 319
QY 327 QPREQVYTL-----PRSDLETNQVSLTCLVKGKGYPDIAVEMESN--GQEPNNKKT 379
Db 320 PRAAEVVAFAATPEWPGSDKRT---LACLIQNPEDISVQWLNHEVQLPDARSTT 374
QY 380 PVLVDSDSFFLYSKLTVDKSRMOGNVPSGVNMHALLNNHYT-QKSLSLSPG 431
Db 375 QPKTKKGGFFVFSRLLEVTRAEWQKDEFICAAVHHAAPSQIVQDAVSYNPG 427

RESULT 70
S00390
Ig gamma chain (clone 36) - chicken (fragment)
N:Alternate names: Ig nu chain
C:Species: Gallus gallus (chicken)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 11-Jan-2000
C:Accession: S00390
R:Parvati, R.; Avital, A.; Lentner, F.; Ziv, E.; Tel-Or, S.; Burestein, Y.; Schechter, I.
EMBO J. 7, 739-744, 1988
A:Title: Chicken immunoglobulin gamma-heavy chains: limited VH gene repertoire, combinatoric
A:Reference number: S00390; MUID:88283642; PMID:3135182
A:Accession: S00390
A:Molecule type: mRNA
A:Residues: 1-504 <PAR>
A:Cross-references: EMBL:X07174
A>Note: This sequence was determined from the differentiated gene
C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: Immunoglobulin

Query Match 15.0%; Score 361; DB 2; Length 504;

Best Local Similarity 28.8%; Pred. No. 9.3e-15;

Matches 95; Conservative 70; Mismatches 129; Indels 42; Gaps 14;

QY 114 QKEEYOL--LVFGLT-ANSDTH-LLOG-OSLTLTLESPPGSSPSVOCSPRGKNIQGGKT 168
DB 192 QSESVELLCTVGFSPASAEVEMLVGCGGLVVASPAVRSGSTYSLSR----- 242
QY 169 LVSQLELDSDGTWTCTVLQONKVEFKIDIVCPAPPEKSCDKHTTCPELLGSPVLEF 228
DB 243 VVWSGDMREKSSYSCRVHHPATNTVVEBHVKGCP-----DGAQSCSPI---QLYAI 291
QY 229 PKKPDITMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKRP--EQVNSTNR 286
DB 292 PPSFGE-LYISLDAKIKRCLVNLPR--SDSSLVYVTRRE--KSGNLRDPVNLDEHNGYTS 347
QY 287 VVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKN 346
DB 348 ASSAVPVSTQDMLSGERFCTYVQHVELRPLSKSVYRNRGPTTPPLTFYFAHPPELISL 407
QY 347 QVSLTCLVKGFPYPSDIAVEM--ESNGQPNNTKTPPVLD-----SDG-SFFLYSKL 395
DB 408 RYTLSTCLVGFPRPRDIEIRMLRDHRAVPATEFVTTALPEERTANGAGGDDPTFVYSKM 467
QY 396 TVDKSRWQGNVPSGVHMEALHNHTOKSLSLSPG 431
DB 468 SVETAKNGTVPACMAVHAELPMRPSQTLQKQAG 503

RESULT 71

MHRB

Ig mu chain C region, secreted form - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 23-Aug-1997
C:Accession: A02164
R.Bernstein, K.E.; Alexander, C.B.; Reddy, E.P.; Mage, R.G.
J.Immunol. 132, 490-495, 1986
A:Title: Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain of V-Ha2 a
A:Reference number: A02164; MUID:84088930; PMID:6418803
A:Contents: a2 allotype
A:Accession: A02164
A:Molecule type: mRNA
A:Residues: 1-458 <BER>
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin F:21-92/Domain: immunoglobulin homology <IMM1>
F:130-202/Domain: immunoglobulin homology <IMM2>
F:242-310/Domain: immunoglobulin homology <IMM3>
F:349-420/Domain: immunoglobulin homology <IMM4>
F:14/Disulfide bonds: Interchain (to light chain) #status predicted
F:28-90,137-200,249-308,356-418/Disulfide bonds: #status predicted
F:46,114,212,261,277,284,445/Binding site: carbohydrate (Asn) (covariant) #status predicted F:219,457/Disulfide bonds: Interchain (to heavy chain) #status predicted
F:296/Disulfide bonds: Interchain (to mu chain in another subunit) #status predicted

Query Match 14.9%; Score 360; DB 1; Length 458;

Best Local Similarity 23.8%; Pred. No. 9.6e-15;

Matches 118; Conservative 76; Mismatches 145; Indels 156; Gaps 22;

QY 21 AATQGNKVVLGKKGDVVELTCTASQ--KKSIOFHW--KNSNOIKILNGSGSLTYGPKSL 76
DB 17 ALTDGGLVVMG-----CLAPDLPSAVTFSSFFKNSSEI----- 50
QY 77 NNRADSRRLMDGNSPELLIKKLL-----IEDDTY-ICEVEDQKEVQL 121
DB 51 -----SKRTV---RTFPVAKRGDKWTATSQVLVPSKDVLTGEVEYLVCKVQSHSNRDLR 102
QY 122 VEGTANSDTHLLQSGSLTLTLESPPGSSPSVOCR-----SPR-- 159

DB 103 V-----SFPVDSLELPNNVSVEFIPPADSPSGSGTRKSRLLICQATGSPKQI 147

QY 160 -----GKNIQGG-----KTLVSQLELDSDGSM--TCTV-----L 187

DB 148 SVSWIRDOQKVESGLTVPEAEETKAGAPAFSSIMLTITRESDLQSLYTCRVDHRGI 207

QY 188 ONOKVEEKIKIDVPCAPPEKSCDKHTTCPELLGSPVFLPFPKPDITMISRTPEVTCV 247

DB 208 FPDKVVSSSSSECSSTTPSP-----GIQFPFAPASADI-FLSKARLLCL 250

QY 248 VVDVSHEDPEVKFNMYVDGVEVNAKTK-----KPREQYNSTRVSVLTVLHQDWLNGK 302

DB 251 VTDLTTVYG-SINISW-----ASHNGKALDTHMNITESHPNATFSAAGCAVCAEDWESGE 304

QY 303 EYKCKVSNKALPAPIEKTISKAKGQPRE--POVYTLPPSRDLT-KNQVSLTCLVKGFPSP 360

DB 305 QFTCTVTIADLPFPLKHTISKREVAKHPAVVYLPPAPRELQVRESATVTCVKGSPSA 364

QY 361 DIAVESNGQPR--ENNTKTPPVLD--SDGSFFLYSKLTVDKSRWQGNVPSGVHMEA 416

DB 365 DVFPVMOQRGQPLPSDDKVTYAPAPPEQAPGLYTHSTLYTTEEDMNGSETFTCVVGHBA 424

QY 417 LHNHTOKSLSLSPG 431

DB 425 LPHWVTEKTVDKSKTG 439

RESULT 72

160082

CD4 receptor - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 23-Jul-1999

C:Accession: 160082

R.Zverev, V.V.; Sidorov, A.V.; Nedospasov, S.A.; Malinshova, V.V.; Udalova, I.A.; Andzha

Vopr. Vitrusol. 40, 100-102, 1995

A:Title: [Nucleotide sequence of two exons of the human T-lymphocyte CD4 receptor gene].

A:Reference number: 160082; MUID:95407135; PMID:7676667

A:Accession: 160082

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-71 <RES>

A:Cross-references: GB:S79267; NID:G1086922; PIDN:AAB35273.1; PID:G1086923

C:Genes:

A:Introns: 17/1

C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

Query Match 14.8%; Score 357; DB 2; Length 71;

Best Local Similarity 98.6%; Pred. No. 1.7e-15;

Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFPHLLVQLALLPATQGNKVVLGKKGDVVELTCTASQKKSIOFHWKNSNOIK 60
DB 1 MNRGVPFPHLLVQLALLPATQGNKVVLGKKGDVVELTCTASQKKSIOFHWKNSNOIK 60
QY 61 ILNGSGSLTYK 71
DB 61 ILNGSGSLTYK 71

RESULT 73

S03961

Ig mu chain C region - house shrew (fragment)

C:Species: Suncus murinus (house shrew)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jan-2000

C:Accession: S03961

R.Ishiguro, H.; Ichihara, Y.; Namikawa, T.; Nagatsu, T.; Kuroawa, Y.

FEBS Lett. 247, 317-322, 1989

A:Title: Nucleotide sequence of Suncus murinus immunoglobulin mu gene and comparison wit

A:Reference number: S03961; MUID:89232144; PMID:2497033

A:Accession: S03961

A:Molecule type: DNA

A:Residues: 1-457 <ISH>

A:Cross-references: EMBL:X13920

C:Genetics:
A:Introns: 106/1; 221/1; 327/1
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:241-309/Domain: immunoglobulin homology <IMM>

Query Match 14.7%; Score 354.5; DB 2; Length 457;
Best Local Similarity 25.5%; Pred. No. 2,1e-14;
Matches 122; Conservative 71; Mismatches 163; Indels 123; Gaps 19;

```
QY 19 LPAATQGNKVVLLGKGGDTVELTCTASQ--KSIQIHMKNKSNQIKLGNQGSFLTKGPKSL 76
DB 17 LPDEFQ-----VTLGLCLARDLPLRPVTFMSKFKKSSSI----- 49
QY 77 NDRADSRSLMDQGNFPLIKKLK-----IEBDDTYI-CEVEDQKEVOQL 121
DB 50 -----SSQNIY--NPEVFTGGKTMATSOVLPESTALIQSTDDYITHTHTTGEKKK 101
QY 122 VEGLTANSPTHLLOQSILTLTLESPGSSPVQCRSPGKNIOGKTLVSQLELDQSG- 180
DB 102 V-----ELQVTEPELPNVISIFV---PFR--NSFGNHPRTSQGLIQASGF 141
QY 181 ---TWTCTVLOQKVEFKIDIVPCPAPBPSCDKT-----HTCP 217
DB 142 SPRTIIVMSWLORGEPVQSLVSTSAVEAPKSGPTTRVISRLTITENEMLSQREFTCQ 201
QY 218 EL-----LG-----GPSVFLPFPKPKDTLMISRTPEVTCVAVDVSHEDPEV 258
DB 202 ALHKRLTQKXVSVCKMGDDSTGTSVFLPPTPMN-IFLTQSAQLTCLVTGLATYD-SL 259
QY 259 KENWYVDGEVYNAKTKPREEQNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIE 318
DB 260 DISMRONGEALQTHVNISESHPNSTFPAKHASVCREMESGEKFTCTVGHQSLDSPDLK 319
QY 319 KTISAKQOPRE-POVYTLPPSRDEL-TRKQVSLTCLVKGFPSDIAYEWESNGQP--EN 374
DB 320 QSLSPPKVAVNDPPVFLPPAQEQLKRESASITCLVKDPSPPVFWQOMHGGQPPVDPK 379
QY 375 NYKTPRPVLD--SDSPFLYSKLTVDKSRMOQGNVFCSSVMEALAHNYTOKSLSLSG 431
DB 380 HYVTNPFPEPNPGLYFVHSLITVSKDMSSGESFCVGHGHALPLSTVEKAVDKTISG 438
```

RESULT 74
S25644
Ig mu chain C region - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Feb-1995 #sequence_revision 30-Jan-1998 #text_change 21-Jan-2000
C:Accession: S25644
R:Paraker, K.; Bugeon, L.; Soullion, J.P.
submitted to the EMBL Data Library, September 1992
A:Reference number: S25644
A:Accession: S25644
A:Molecule type: mRNA
A:Residues: 1-343 <PAR>
A:Cross-References: EMBL:X68312; NID:G56461; PIDN:CAA48392.1; PID:G818025
A:Experimental source: spleen
C:Genetics:
A:Map position: 6
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:234-305/Domain: immunoglobulin homology <IMM>

Query Match 14.6%; Score 353; DB 2; Length 343;
Best Local Similarity 30.0%; Pred. No. 1.8e-14;
Matches 87; Conservative 66; Mismatches 107; Indels 30; Gaps 10;

```
QY 152 SVQCRSPGKNIOGKTLVSQLELDQSGTWTCTVLOQKVEFKIDIVPCPAPBPSCD 211
DB 55 TVEAGSRPQYKXVSTLITSDMLNINVTCTRV--DHRGLTFKXVNSSTCAASPST-- 110
QY 212 KTHTCPELLGSPVFLPFPKPKDTLMISRTPEVTCVAVDVSHEDPEVKFNWYDGEVHN 271
```

```
DB 111 -----DILAFP-----IPSPFAD-IFLTKSAKLSCLVTNLATYD-TLNISSSKSGEPLE 158
QY 272 AKTPREEQNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIKTSKAGOPRE- 330
DB 159 TWTIMESHPRGTFSAAGVAVSCMEDNKRKEFCYTHDLPSPQKRFISK----PNEV 214
QY 331 ---POVYTLPPSRDEL-TRKQVSLTCLVKGFPSDIAYEWESNGQP--ENNYKTPPV 383
DB 215 AKHPAVAYLLPPAREGLILRESATVTLVKGFSPADIFVQWLORGQGLSSDKYTSAPMP 274
QY 384 D--SDGSFPLYSKLTVDKSRMOQGNVFCSSVMEALAHNYTOKSLSLSPG 431
DB 275 EPGAPGLYFTHSILTVTEEMNSGETYTCVGHGHALPLHMYTERTVDKSTG 324
```

RESULT 75
MHNY
Ig mu chain C region - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Aug-1996
C:Accession: A02168
R:McGuire, K.L.; Duncan, W.R.; Tucker, P.W.
Nucleic Acids Res. 13, 5611-5628, 1985
A:Title: Phylogenetic conservation of immunoglobulin heavy chains: direct comparison of
A:Reference number: A02168; MUID:85297761; PMID:2994005
A:Accession: A02168
A:Molecule type: DNA
A:Residues: 1-454 <MCG>
A>Note: the sequence was determined from the germline gene
C:Genetics:
A:Introns: 105/3; 218/3; 324/3
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 161
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F:128-200/Domain: immunoglobulin homology <IMM2>
F:238-306/Domain: immunoglobulin homology <IMM3>
F:345-416/Domain: immunoglobulin homology <IMM4>
F:45,113,139,192,210,238,257,280,326,441/Binding site: carbohydrate (Asn) (covalent) #acc:
F:135-198,245-304,352-414/Disulfide bonds: #status predicted

Query Match 14.5%; Score 349; DB 1; Length 454;
Best Local Similarity 24.6%; Pred. No. 4.5e-14;
Matches 113; Conservative 82; Mismatches 154; Indels 110; Gaps 19;

```
QY 37 VELTCTASQ--KSIQIHMKNKSNQIKLGNQ--SFLTGPSPKLNDRADSRSLMDQGNF 92
DB 23 VAMQCLARDLPSISISFSMNYQNKSEV--NQGVATFPTL-----RMGEKYAATSQVFL 73
QY 93 PLITKLIKIDSDPY-ICEVE--DQKEVQLVTPGLTANSPTHLLOQSILTLTLESPGS 149
DB 74 P--PKSVLEGSDEYLVCVHAGNTKNDLVLPVPGV-----EMNPNV 113
QY 150 SPVSQCR-----SPR-----SKNIQCG----- 166
DB 114 SVFPPSRDAPSGAPRKRSLFCEASNSPKQITVSWLRDQPKVSGTTERPVYEDGSG 173
QY 167 -----KTLVSQLELDQSGTWTCTVLOQKVEFKIDIVPCPAPBPSCDKTHTCPEL 219
DB 174 PRYKIVSTLTLTESDMLNINVTCTRV--DHRGLTFKXVNSSTCAASPST----- 221
QY 220 LGGSVSLPFPKPKDT--LMSRPEVTCVAVDVSHEDPEVKFNWYDGEVYNAKTPRE 278
DB 222 ---DIDAFIPBPFGVIFLTKSATYTLCLVTNLATYD-TLNISSRSGBPLETKTKL 276
QY 279 EGVNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIKTSKAGOPR-EPVYTL 337
DB 277 SHPMTSALGEANVCBEDDSGKEFVCTYTHDLPSPQKRFISKPGKMNKTPPAVYQOP 336
QY 338 PSRDEL-TRKQVSLTCLVKGFPSDIAYEWESNGQP--ENNYKTPPVLDSDGS--FFLY 392
```

Db 337 LAEQLILRESATVTCLVKGFSPADIFVQWLQKQPLSQDKVTSAPMBEQAPHLVFTN 396
QY 393 SULTVKSRRMOQGNVSCSVMEALHNHTQKSLSPG 431
Db 397 SVLTVEEENSGEYTCVVGHEALPHMVTERTVDRSTG 435

RESULT 76

C31933
Ig mu chain C region - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 31-Mar-1990 #sequence_revision 30-Jun-1992 #text_change 20-Sep-1999
C:Accession: C31933
R:Schwager, J.; Mikorjak, C.A.; Steiner, L.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 2245-2249, 1988
A:Title: Amino acid sequence of heavy chain from Xenopus laevis IgM deduced from cDNA se
A:Reference number: A94192; MUID:88176921; PMID:2451244
A:Accession: C31933
A:Molecule type: mRNA
A:Residues: 1-453 <SCH>
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: Immunoglobulin

Query Match 14.3%; Score 345.5; DB 2; Length 453;
Best Local Similarity 25.8%; Pred. No. 7.3e-14;
Matches 119; Conservative 77; Mismatches 189; Indels 77; Gaps 18;

QY 31 GKKGDTVELTCTASQ--KSIQPHMKNSQIKLQSGF---LTGSPSKLNDRADSR 84
Db 15 GESMDVITICLAKDELPERISFTMGDKNNAASYTLKSKYKPMQSGTYSASSQVNVAS 74
QY 85 SLWDQGNFLILNKLIEISDTYICGEVDQKEVQ-----LLV 122
Db 75 AVMD--NIEQFYNAHLDT--IKVELKQDPVKEKRVVISIHPPSKDALALNESLFI 129
QY 123 FGLTAN-SDTHLQGSSTLTLESPPGSSPVQCRSP---RGNLQGGKTLASVQLELD 178
Db 130 VCLATNFTPTHT-----VTKMLKNGQTTGCVARVEEPVDKKRGVATYSTLTREMDL 184
QY 179 SCMTCTVLQONQKQKFEKIDIVCPAPPEPKSCDKTHTCEBLGSPS--VLEPPPKDYL 236
Db 185 DTLVSCV-----EHAESGSLQEKMSKSLMCDTPIPTISIQVITTPPS--LES1 232
QY 237 MISRTGEVTVVVDVSHEDBEVKNVY--VDGVEVNAHAKKPREQVNS--TYRVSVTLV 293
Db 233 FEKKSATLTCLVSNMANSDELRSISWPKSGTQELPLKTELGDALINDRTYSVKGTTV 292
QY 294 LHDMLNGKVKYKCNKALPAPIEKTIISKAKQPREPVYTLPPSRDELTKNQ--VSLTC 352
Db 293 CADEMNDK-FVCKVEHTELASKKEVFLPREKEGVNTPSVYVPPLEBELSKRETATLTLC 351
QY 353 LVKGFPSDIAVEW--ESNGQPENNYKTPP-----PVLSDGSPFLYSKLTVDKSRWQ 404
Db 352 LVVGFPSSEIFVYKMLHKNZAVPKONYINTSINDELLPKQKSGKFLYSLHTIDIKMDA 411
QY 405 GNVEGCVNHEALHNHYTQKSLSPG-----QLDETC 438
Db 412 GDSFSCVVGHESLPLQLTQRSIDKSSGKPTNVNVSVLSDTC 453

RESULT 77

B30503
Ig gamma-2a chain C region (B5.7A12) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1989 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: B30503
R:Gilmore, G.L.; Bard, J.A.; Birstein, B.K.
J. Immunol. 141, 1754-1761, 1988
A:Title: DNA rearrangements affecting both variable and constant regions of Ig H chain g
A:Reference number: A30503; MUID:88315788; PMID:2842402
A:Accession: B30503
A:Molecule type: mRNA
A:Residues: 1-112 <GIL>

A:Cross-references: GB:M21925
A:Experimental source: myeloma cell line MPC11
A>Note: the authors translated the codon GAG for residue 11 as Ser
C:Genetics:
A:Introns: 100/3
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:1-70/Domain: immunoglobulin homology <IMM>

Query Match 13.7%; Score 331.5; DB 2; Length 112;
Best Local Similarity 50.0%; Pred. No. 1e-13;
Matches 66; Conservative 17; Mismatches 22; Indels 27; Gaps 2;

QY 239 STPEVTCVVVDVSHEDBEVKFNWYVDQGVFNAAKTKREDOVNSYTRVSVTLVLDQDW 298
Db 1 SLTPRYTCVVVDVSHEDDPVQISWFTVNNVTAQTQTHREDYNTIRVSTLPIQHDW 60
QY 239 LNKGEYKCKVSNKALPAPIEKTIISKAKQPREPVYTLPPSRDELTKNQVSLTCLVKGFY 358
Db 61 MSGKEFKCKVNKKDLPAPIERTISKGE-----SCSLTAMGLGW- 100
QY 359 PSDIAVEWESNG 370
Db 101 -----WTSNG 105

RESULT 78

I50731
Ig heavy chain - nurse shark
C:Species: Ginglymostoma cirratum (nurse shark)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I50731
R:Yazquez, M.; Mizuki, N.; Flajnik, M.F.; McKimsey, E.C.; Kasahara, M.
Mol. Immunol. 29, 1157-1158, 1992
A:Title: Nucleotide sequence of a nurse shark immunoglobulin heavy chain cDNA clone.
A:Reference number: I50731; MUID:92357056; PMID:1495502
A:Accession: I50731
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-577 <VAZ>
A:Cross-references: GB:M92851; NID:9213264; PIDN:AAA50817.1; PID:9213265
C:Genetics:
A:Gene: IGH
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:1469-539/Domain: Immunoglobulin homology <IMM>

Query Match 13.5%; Score 325.5; DB 2; Length 577;
Best Local Similarity 24.3%; Pred. No. 1.6e-12;
Matches 142; Conservative 83; Mismatches 165; Indels 195; Gaps 28;

QY 10 LILVLTQALLPRAIQGNRV-----LKKKGDVELTCTASQKSIQPHMKNSQ--IKIL 62
Db 6 IFLSLILALLPCVQSEITLQPEAEYGHGGSLSLTC-----KTRGFNLGSSSMYMIQV 60
QY 63 GNGQ-----SFLTGPSKLANDRADSRSLIMQGN--PPLITKNIKIEDSDTYICEV 111
Db 61 PEQGLEMTVYIYSSSMNYPATIKDRFTAAC--DTSNNITALEKRSVKIDTALTYC-- 115
QY 112 EDQKEEVLVFGLTANSDTHLQGSSTLTLESPPGSSPV-----QCRS----- 157
Db 116 -----TRMSGVEYLIGHSGYWGQGTWVTVTATP--SSPTLYGLVSSCQGGNIDGSVIY 168
QY 158 -----PRGKNIQGGKTLVSQI----- 174
Db 169 GCLANDVSPDVASVTWKHGLQITTVGTQVPSVRNKKGTYLS--SQLALIESDAECDQIS 227
QY 175 -ELDQSG----- 185
Db 228 CEVRHSIGSDKSTGMCPDGPFPALTIVSSSEIESRKAIIIVCSISDPHSKISVYTW--- 284
QY 186 VLQONKQVFEKIDIVP--CPA-----PEPKSCDK--THTC-----PEL-----LGG 222
Db 285 -LKNGRSVDSGIFTSFVCEANGNFSVTSRLAVPYAEMWDBRAVYITQGVKXVIGSQWNITG 343

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1474 <MIL>
A:Cross-references: EMBL:U12455; NID:g529949; PIDN:AAC59688.1; PID:g529950
C:Gene: 19M
C:Superfamily: immunoglobulin C region; immunoglobulin homology

Query Match 12.8%; Score 308.5; DB 2; Length 474;
Best Local Similarity 24.2%; Pred. No. 1.4e-11;
Matches 116; Conservative 62; Mismatches 171; Indels 131; Gaps 20;

```
QY 31 GKKGDPVLELTCTAS--QKSIQPFHMKNSQIKILGNQSGFLTPGSKINDRADSRSLMD 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 50 GTSQDVVALGCLATGSLTPSKMTDSTDKEL-----TPFRKYSVLNGETYSSTIS--- 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 89 QGNFPLIIKLIKIEDS-----TYICEVEDQKEVOLLVFGLTANSDFHILQOSLTLTL 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 102 -----QLSLPTSDMNSGKAFCEAKHPQGDVLT-----HLL----- 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 ESPSGSPVQCSPPGSKNIQSGKTLVSQLELDQSGTWTCTVLQ--NOKKVEPK----- 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 -TPFPVPAVLLNLP-----SLBEFAQNHATATLVCSRGSFSPKTHFEKMRGNT 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 197 -----IDIVCPAPPEPK-----SCDKTH-----TC 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 KLDGVTNIPATYDEKLLYSASSLLVTTEKDKKSAEFACEFVHKTSVLKNIITYTSREK 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 217 PELIGSPVFLFPPPKEDTLMTSRPEVTC---VVVDVSHEDPEVFNWYVDGVEVHNK 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 QETV---KVIEPPTNEEQ--FVKKTATLTCRRLALVSTG---DVSMTSSSGKPL--AA 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 274 TPRREQYNSTRKVSVLTVLHODMLNGEYKCKYSN--KALPAIEKTISSAKQPREPQ 332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 291 GAREFHGEGKIVAVRSVYLEKMTGTEYKCTIVHLSFPTPIKTYKQIATKIRPS 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 333 VYTLPPSRDE--LTKNQVSLTCLVKGFPSPDIAMVEM--ESNQGPENNYTPTPVLDSDS 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 VFLPABSTENSTREDEVTLTFCVDFKDFSPKDIYLSLADSDYDKHVTVDLIPSHDA 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 389 FFLYSKLTVDKSKMOQGNVPSGCVMEHA-----LHNHYTQKSLSPGLQDTEK 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 YGVYSKYTITSSSDMNSGTMYSQAVNHETAPLPVSVITRTTDSSTGKALIVNFTLNPDIK 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 82

S21461
T-cell surface glycoprotein CD4 (allele 1) - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000
C:Accession: I47131; S21461
R:Guatafeson, K.; Germana, S.; Sundt, T. M.
J. Immunol. 151, 1365-1370, 1993
A:Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine C
A:Reference number: I47131; MUID:93329116; PMID:8335933
A:Accession: I47131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-99 <GU2>
A:Cross-references: EMBL:X65629; NID:g1928; PIDN:CAA46583.1; PID:S388232
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: glycoprotein; T-cell
F:3-81/Domain: immunoglobulin homology <IMM>

Query Match 12.7%; Score 305.5; DB 2; Length 99;
Best Local Similarity 60.2%; Pred. No. 3.4e-12;
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

```
QY 32 KKGDPVLELTCTASQKSIQPFHMKNSQIKILGNQSGFL--TKGPKSKINDRADSRSLMDG 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 KAGDLELPCHSQKKNLPFNWNSQTKILGSGFMHTASVETLSRLDSKKNWMDHG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 91 NFPLIIKLIKIEDSDTYICEVEDQKEVOLLVFGLTAN 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 61 SFPLIIKLIKIEVDSDTYICEVEDQKEVOLLVFGLTAN 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 83

A46507
Ig alpha chain - chicken
C:Species: Gallus gallus (chicken)
C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C:Accession: A46507
R:Maniaka, A.
J. Immunol. 149, 855-861, 1992
A:Title: Chicken IGA H chains. Implications concerning the evolution of H chain genes.
A:Reference number: A46507; MUID:92340889; PMID:1634774
A:Accession: A46507
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-585 <MAN>
A:Cross-references: GB:S40610; NID:g251907; PID:g251908
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBI:109906, NCBI:109907)
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match 12.5%; Score 301.5; DB 2; Length 585;
Best Local Similarity 26.8%; Pred. No. 4.7e-11;
Matches 114; Conservative 60; Mismatches 184; Indels 67; Gaps 22;

```
QY 19 LPATQGNKVVLLGKKGDPVLELTCTASQKSIQPFHMKNSQIKILGNQSGFLTPGSKIND 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197 LPVATTTGTYSL-----LTALTVPREQ-----LQGN--FVCRAGHAATG 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 79 RADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEVOL--LVFGL--TANSDTHILQ 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 -ADVAKETIGDGVCPITFSKYTLISDPDQ---EDFERVLLVCLVEGLPSKGAALQWLQ 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 -QGSLLTLLESPSGSPVQCSPPGSKNIQSGKTLVSQLELDQSGTWTCTVLQ--NOKKVE 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 290 DNEENT---PAPSDSGSGDCTESGVTQMSRVNVRKSWEGAQPGCRTHGALK-- 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 FKIDIVCPAPPEPKSCDKTHTCPELIGSPVFLFPPPKEDTLMTSRPEVTCVVVDVSH 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 343 -----EPYATATSTDCAT---PQL-----QVSLPPTLEF--LVSHNKTATVTCVSNAAA 389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 DPEVFNWYVD--GVEVHNKTKPREQYNSTRKVSVLTVLHODMLNGEYKCKYSNK 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 390 D-GVGSVMSRSRSGGLDV---SQTEDRQADGRYTVRSFLRVCABEMNGGERTFGCSVRE 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 312 ALPAIEKTISSAKQKQ--REPOVYTLPPSRDELT--KNQVSLTCLVKGFPSPDIAMVEMESN 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 445 GV-VVAESIRKETDTPPHAPSVVFPFPPABELSLQETATTLTCMASSFLPSSILLTWTOQ 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 GQP--ENNYKTTPTPLVLDSDGSFF--LYSKLTVDKSKMOQGNVPSGCVMEHALHNHYTQKSL 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 504 NQPIPPQVYLLFGP--EKDGFYSLSKLVSVEMQGDVFGCVGHGDIPLANTHHSI 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 SLSPG 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 562 DKNAG 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 84

S60266
novel antigen receptor precursor - nurse shark
C:Species: Ginglymosoma cirratum (nurse shark)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 21-Jul-2000
C:Accession: S60266
R:Greenberg, A.S.; Avila, D.; Hughes, M.; Hughes, A.; McKinney, E.C.; Flajnik, M.F.
Nature 374, 168-173, 1995
A:Title: A new antigen receptor gene family that undergoes rearrangement and extensive B
A:Reference number: S60266; MUID:95183140; PMID:7877689
A:Accession: S60266
A:Status: preliminary; nucleic acid sequence not shown

QY 217 PELLGGBSVFLPFPKXKDTLMTSRTPVYTCVVDVSHEDP-EVKFNNYVDGVEVHNATK 275
Db 130 PVQODIAIRVITPSPVD-IFISKATLTICRVSNNVNDAGLEVSW-WKEKGKLETLGK 187
QY 276 PREOVSNTYRVVSVLTVLHODMLNGEKYKCVSNKMLPAIEKTTISKAK-GQPREQVY 334
Db 188 -RVLONSGLTYDGVATVACASEMDGDGGVCCVNNPDLTFPMEKMKTKTASNSNRPVY 246
QY 335 TLPPSRDELTKNQ-VSLTCLVAGFPSPDIAVEMESNGQ--PENNYKTTTPVLVS--DGSF 389
Db 247 VFPPPEPTEQLNGNQRSLVTCMAAGFNRPPLFLFRWMRNGBPPLQDSQSVTSAPMAENPENESY 306
QY 390 FLYSKLTVDKSRMOQGNPFSCVMEEDALHNHYTKSLSLSG 431
Db 307 VAYSVYGVAAEEGAGNNYVTLCLVGEHALPLQIACQSVDRASG 348

RESULT 89
A46532
ig mu chain C region - axolotl (fragment)
C:Species: Ambystoma mexicanum (axolotl)
C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A46532
R:Fellah, J.S.; Miles, M.V.; Charlemagne, J.; Schwager, J.
Eur. J. Immunol. 22, 2595-2601, 1992
A:Title: Evolution of vertebrate TGM: complete amino acid sequence of the constant region
A:Reference number: A46532; MUID:93011455; PMID:1382992
A:Accession: A46532
A:Status: Preliminary
A:Molecule type: nucleic acid
A:Residues: 1-454 <FEEL>
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:115554)
C:Superfamily: immunoglobulin C region; immunoglobulin homology
;Keywords: immunoglobulin

Query Match	12.3%	Score 296;	DB 2;	length 454;
Best Local Similarity	23.1%;	Pred. No. 7.5e-11;		
Matches 111;	Conservative 65;	Mismatches 190;	Indels 114;	Gaps 17
Qy	21	AATQGRKV----	LGKKDPVELNCTAS--	OKKSIOCFMKNQNIKI---LGNQSFLLTK 71
Db	1	AAAGASVYPLISCGASLDPVIGCLAGPLPDSVTEDMDMDKNNASISAGVAKLPSTTIG 60		
Qy	72	GSFKLMDRADSRSILMDQGNFPLIKLKIEDSTTYICEYEDQKEVQLLVFGLTANST 131		
Db	61	GLYSATRKVPSBSM-----	KNRDPYCKV-----	T 87
Qy	132	HLHGGSLTLTLESPG--SSPSVQCRSP-----		R 159
Db	88	HBELQGPLTKKQVFSQVQRI SKPTYLTHAPAEEDIINNATTVCICRPFHPQIPISIKMK 147		
Qy	160	GKNIQGGKTLVSQLELDQSGTWCTVLQNGKVEFKIDIV-PCPAE-----		PKSCD 211
Db	148	GKQVTSG--IYEEBVDADTAGNFDVTSILNIEPEMDMDTVYSCVDQTSKPFMNINMS 205		
Qy	212	KHTHCPELLGGSVFLPFPKPK-DLYMSRPETVCCVVUV-----		SHEDPEVK 260
Db	206	KSNLCDAQVGPRKATATFAPAPFEEMFESKANATCTVYNNGITEGNTIWSEDTNEVL 265		
Qy	261	NMYDGEVHNAAKTPREBYNNSYTVRVSVLVTHDLMGKEVKKCVSNKALPAPIEKT 320		
Db	266	KTEITPIPIFD-----	NATLSVWGIVTCADQAMDHNFVCKVLTHQDLAEORVLS 315	
Qy	321	ISKAKG-OPREQVYTLPRPSDELT-KNOVSLTLVGVGFPSDIAVEMESNG--PENNY 376		
Db	316	LQPKNGNQRRKESVITYPPSEBELAKETATVCLMMGYNHCDLFPYRLMESQOLOKODT 375		
Qy	377	KTTTPVLDSD-----	GSFFLYSKTLVDKSRWQGNVESCSVMEALHNHTYOKSLSLSPG 431	
Db	376	VNTKQLEVEDPTTGQKSCFMYSMLKIPAAQDANTATNTVCLVGHDAFLPLOTQKSIDRSG 435		

EMBO J. 8, 4041-4047, 1989
A>Title: The 19A heavy-chain gene family in rabbit: cloning and sequence analysis of 13
A:Reference number: S09264; MUID:90076124; PMID:2512120
A:Accession: S09276
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-338 <BUR>
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:228-300/Domain: immunoglobulin homology <IMM>

Query Match 12.1%; Score 292; DB 2; Length 338;
Best Local Similarity 29.2%; Pred. No. 9.4e-11;
Matches 92; Conservative 50; Mismatches 133; Indels 40; Gaps 15;

146 PPGSSPVQCRSPRKNIGGKTLVSQLELSDSGTWTC-----TVLQNOCK---V 193
16 PPG--PLSVSWTVNGEN---SAVFPQAQSTSGPYACSELIIPVQCIGQKSACHV 70
Qy 194 EFKIDI---VPCPAPEPKSCDKHTKCP-----ELIGG--PSVFLPPPKDTLMISRTPE 243
71 EYNSTINSLPVPFPDCCPANCCTCPSSSSRNLIJSGQPSISLGRPLGD--LILGRDAS 129
Db 244 VTCVVVDVSHEDPEVKFWMYDGVENNAKTKPRBEQYNSTYRVVSVLTVLHQDWLNGKE 303
130 LTCITSLGKLNEDAV-FTW--EPTNGNBPVQORACORDLSCGVSVSVLPSSAETWKARTE 186
Qy 304 YKCKVSNKALPAP-PEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VQLTCLVKGFPYPSD 361
187 FCTYVTHEIDSGSLTATISKGVTP--PQVHLPPPEELALNEQVTLTCLVKGFSKXD 244
Qy 362 IAVENESNGQ--PENNY---KTPPEVLDSGSPFLYSKLTVDKSRMOGQNVFSGSVMEHA 416
245 VLVSRRHOGQVEPDSFLVWKSMPSSQDKATYATISLRLVPAEDMNGDGYSCVGHG 304
Db 417 LHNHTQKSLSPG 431
305 LAEHTORTIDRLAG 319

RESULT 96
S09266
Ig alpha chain C region - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999
C:Accession: S09266
EMBO J. 8, 4041-4047, 1989
A>Title: The 19A heavy-chain gene family in rabbit: cloning and sequence analysis of 13
A:Reference number: S09264; MUID:90076124; PMID:2512120
A:Accession: S09266
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-352 <BUR>
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:241-314/Domain: immunoglobulin homology <IMM>

Query Match 12.1%; Score 292; DB 2; Length 352;
Best Local Similarity 31.2%; Pred. No. 9.8e-11;
Matches 95; Conservative 41; Mismatches 118; Indels 50; Gaps 16;

Qy 139 LTLLESPGSSPVQCRSPRKNIGGKTLVSQLELSDSGTWTCVTLQNOCKVEFKID 198
Db 69 LSTLEQCP-EDHNVCAVEH--NYDEGONLTVLXPECKDRPS-----D 109
Qy 199 IVPCAPPEPKSCDKHTKCP-----ELIGG--PSVFLPPPKDTLMISRTPEVTCVVVDVSHEDPE- 257
110 PTPCPCP-PITC-----GEPSSLQRPDID--LLESNASTLTCTLSGL--KDEG 155
Qy 258 VKENNY-VDGVENNAKTKPRBEQYN-STYRVSVLTVLHQDWLNGKEVKCKVSNKALP- 314
Db 156 AVFTNPTNGNEFVQOST---QSYPCGCVSVSVLPQCAEPWNAETFTCTVTHPELEG 211

Qy 315 APIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYPSDIAMVENSGQ-- 371
Db 212 GSLTATISRGSILP--PQVHLPPPEELALNLTCLVGRGSPDVLVSWTHNGTLV 269
Qy 372 -PENNY---KTPPEVLDSGSPFLYSKLTVDKSRMOGQNVFSGSVMEHALHNHTQKSL 427
Db 270 VPKXSLFWKRLPEPGEPPTVAVTSILRLVPAEDMNGDSYCVGHGLAEHTORTID 329
Qy 428 LSPG 431
Db 330 RLAG 333

RESULT 97
A34891
Ig heavy chain precursor V region - ladyfish
C:Species: Elops saurus (ladyfish)
C:Date: 20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change 21-Jul-2000
C:Accession: A34891
R:Amemiya, C.T.; Litman, G.W.
Proc. Natl. Acad. Sci. U.S.A. 87, 811-815, 1990
A>Title: Complete nucleotide sequence of an immunoglobulin heavy-chain gene and analysis
A:Reference number: A34891; MUID:50138916; PMID:2105490
A:Accession: A34891
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-568 <ANB>
C:Cross-references: GB:M26182; NID:g213134; PIDN:AA49238.1; PID:g213135
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:33-116/Domain: immunoglobulin homology <IMM>

Query Match 12.0%; Score 290; DB 2; Length 568;
Best Local Similarity 22.3%; Pred. No. 2.3e-10;
Matches 92; Conservative 78; Mismatches 152; Indels 90; Gaps 15;

Qy 65 QGSFLLTGPKSLNDRADRSRLMDQGNPILI-----KNLKIED 103
Db 207 EGDFKSAEFCCTTELGKKT-----PVVLPKPEPKPPRQPVLSIMTPSQBELTLNK 259
Qy 104 SDTYICEVDQKEVOLLVGLTRANSDTHLLOGOSLTLTESPPGSSPVQCRSPRKN 163
Db 260 TATPACIATP-----FYPKGSHPKWLKRDGEVTP-----DGIATLTCCOKGDKSP 304
Qy 164 QGKTLVSQLELSD--SGTWTCVTLQNOCKVE--FKIDIVPCPAPEPKSCDKHTKPEL 220
Db 305 TASSFLQASQWKRLDGTTCQPIQGEITEQVTKSSAEC--SPEAQIDAK----- 355
Qy 221 GGPVFLPPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKENWYDVE--VHNAKTKPRE 278
Db 356 -----ISPPTPEE-LPQQTRTLTC-----KITGDVGVANVTWVGSEVRV 396
Qy 279 EQVNSTYRVVSVLTVLHQDWLNGKEVKCKVSNKALPAPIEKTISK-AKGQPREQVYTL 337
Db 397 GQFDE-QKMSKSLIDIEEWNKRNREYCKVHSLPRLTSYRRECGKQKQSTVILA 455
Qy 338 PSRDELTKNOVSLTCLVKGFPYPSDIAMVENSGQPENNYKTPPEVLDSGSPFLYSKLT 397
Db 456 PA-EQRRLSTVTLICAKDYPPEQVLSLWLDQPVETDVPTTEVTVTGETISVFSQULT 514
Qy 398 DKSRMOGQNVFSGSVMEHAL-----HNHTQKSLSPGLDDETC 438
Db 515 PASDMDGGVYVSCAVHETWESVVKTIVRTDSVSKPPTVSLDLNVPTC 566

RESULT 98
S09265
Ig alpha chain C region - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999
C:Accession: S09265
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.

Search completed: August 3, 2004, 13:14:55
Job time : 16.4339 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:44 ; Search time 6.23487 Seconds

(without alignments)
3791.557 Million cell updates/sec

Title: SEQ4
Perfect score: 2414
Sequence: 1 NMRGVFRRLLVLQLALP.....DETCAAGDGEJDLMTTDP 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1227.5	50.8	330	1	GCI_HUMAN
2	1133	46.9	290	1	GC3_HUMAN
3	1130.5	46.8	326	1	GC2_HUMAN
4	1119	46.4	327	1	GC4_HUMAN
5	1033	42.4	458	1	CD4_HUMAN
6	999	41.4	458	1	CD4_PANTR
7	955.5	39.6	398	1	GC3M_MOUSE
8	932.5	38.6	393	1	GC1M_MOUSE
9	916.5	38.0	458	1	CD4_MACEA
10	910	37.7	458	1	CD4_MACEU
11	906.5	37.6	399	1	GCAN_MOUSE
12	904	37.4	458	1	CD4_MACMU
13	903	37.2	458	1	CD4_MACNE
14	899	37.2	323	1	GC_RABIT
15	894.5	37.1	329	1	GC2_CAVPO
16	885	36.7	458	1	CD4_CERAB
17	877.5	36.4	405	1	GCER_MOUSE
18	857.5	35.5	329	1	GC3_MOUSE
19	833	34.5	333	1	GC8_RAT
20	823.5	34.1	326	1	GCI_RAT
21	823	34.1	329	1	GCC_RAT
22	818.5	33.9	324	1	GCI_MOUSE
23	805.5	33.4	330	1	GCAN_MOUSE
24	801	33.2	335	1	GCAB_MOUSE
25	792.5	32.8	322	1	GCA_RAT
26	784	32.5	397	1	CD4_CERTO
27	783	32.4	397	1	CD4_ERYPB
28	781.5	32.4	336	1	GC8_MOUSE
29	737.5	30.6	457	1	CD4_SATSC
30	610.5	25.3	463	1	CD4_CANPA
31	601.5	24.9	459	1	CD4_RABIT
32	495	20.5	457	1	CD4_RAT
33	475	19.7	457	1	CD4_MOUSE

34	388	16.1	421	1	EPC_MOUSE
35	384.5	15.9	454	1	MUC_HUMAN
36	373.5	15.5	476	1	MUCN_MOUSE
37	370.5	15.3	429	1	EPC_RAT
38	369.5	15.3	479	1	MDCM_RABIT
39	369	15.3	450	1	MUC_CANFA
40	365.5	15.1	391	1	MUCB_HUMAN
41	362	15.0	455	1	MUC_MOUSE
42	361.5	15.0	428	1	EPC_HUMAN
43	360	14.9	458	1	MUC_RABIT
44	354.5	14.7	457	1	MUC_SUNMU
45	344	14.3	454	1	MUC_MESAU
46	342	14.2	438	1	HVC2_HETPR
47	310.5	12.9	461	1	MUC_CHICK
48	295.5	12.2	461	1	HVCN_HETPR
49	294.5	12.2	340	1	ALC2_HUMAN
50	294	12.2	438	1	HVCS_HETPR
51	293	12.1	299	1	ALC_RABIT
52	286.5	11.9	370	1	HVCI_HETPR
53	281	11.6	393	1	HVCI_HETPR
54	279	11.6	353	1	ALCI_GORGO
55	277.5	11.5	353	1	ALCI_HUMAN
56	275	11.4	481	1	MDCM_ICTPU
57	264.5	11.0	344	1	ALC_MOUSE
58	198	8.2	513	1	SHS1_MOUSE
59	188.5	7.8	506	1	SHS1_BOVIN
60	188.5	7.8	509	1	SHS1_RAT
61	177.5	7.4	105	1	LAC1_MOUSE
62	176.5	7.3	739	1	SRE2_HUMAN
63	175	7.2	387	1	SRE2_HUMAN
64	172	7.1	104	1	LAC2_RAT
65	168.5	7.0	4391	1	PCBM_HUMAN
66	167	6.9	847	1	CD22_HUMAN
67	165	6.8	398	1	SRE1_HUMAN
68	164	6.8	6632	1	UN89_CABEL
69	163	6.8	104	1	LAC3_MOUSE
70	163	6.8	3707	1	PCBM_MOUSE
71	162.5	6.7	213	1	ILL1_HUMAN
72	161	6.7	104	1	LAC2_MOUSE
73	161	6.7	258	1	HB2D_PIG
74	159.5	6.6	1197	1	CAM1_BRARE
75	159	6.6	105	1	LAC_HUMAN
76	157	6.5	739	1	VCA1_HUMAN
77	156.5	6.5	503	1	SHS1_HUMAN
78	156	6.5	105	1	LAC_PIG
79	153.5	6.4	105	1	LAC5_MUSSP
80	153	6.3	106	1	LAC_RABIT
81	153	6.3	106	1	KACB_RABIT
82	151.5	6.3	383	1	DPC_HUMAN
83	150	6.2	106	1	KACA_RAT
84	149.5	6.2	261	1	HB2C_PIG
85	149	6.2	348	1	KILO_RAT
86	148	6.1	106	1	KAC_HUMAN
87	148	6.1	1369	1	NFAS_CHICK
88	147	6.1	338	1	LAMP_HUMAN
89	147	6.1	1240	1	NFAS_MOUSE
90	146	6.0	104	1	LAC1_RAT
91	146	6.0	1259	1	CAME_RAT
92	145.5	6.0	103	1	LAC_CHICK
93	145	6.0	106	1	KACB_RAT
94	144.5	6.0	333	1	AMAL_DROME
95	144.5	6.0	739	1	VCA1_MOUSE
96	144	6.0	338	1	LAMP_RAT
97	143.5	5.9	1240	1	NFAS_RAT
98	143	5.9	106	1	KAC_MOUSE
99	143	5.9	1260	1	CAM1_MOUSE
100	142.5	5.9	1240	1	NFAS_HUMAN
101	140	5.8	268	1	HB2X_HUMAN
102	140	5.8	273	1	2DOB_HUMAN
103	140	5.8	1242	1	NPHN_MOUSE
104	139.5	5.8	702	1	CEAS_HUMAN
105	139	5.8	273	1	2DOB_PANTR
106	136	5.6	2012	1	DSCA_HUMAN

P06336	mus musculus
P01871	homo sapien
P01873	mus musculus
P01855	rattus norv
P04221	oryctolagus
P01874	canis fam1
P04220	homo sapien
P01872	mus musculus
P01854	homo sapien
P03988	oryctolagus
P20768	sinus murt
P06337	meocricetu
P23085	heterodontu
P01875	gallus gall
P23088	heterodontu
P01877	homo sapien
P23087	heterodontu
P01879	oryctolagus
P23084	heterodontu
P23086	heterodontu
P20758	gorilla gor
P01876	homo sapien
P23735	ictalurus p
P01878	mus musculus
P97797	m protei-n-t
O46631	bos taurus
P97710	r protei-n-t
P01843	mus musculus
P23954	rattus norv
O991w8	homo sapien
P20767	rattus norv
P98160	homo sapien
P20273	homo sapien
O00241	homo sapien
O01761	caenorhabd
P01845	mus musculus
O05793	mus musculus
P15814	homo sapien
P01844	mus musculus
P15983	sus scrofa
O90478	brachydanio
P01842	homo sapien
P19320	homo sapien
P78324	h protei-n-t
P01846	sus scrofa
P20765	mus spretus
P01847	oryctolagus
P01839	oryctolagus
P01880	homo sapien
P01836	rattus norv
P15982	sus scrofa
O920j8	rattus norv
P01834	homo sapien
O42414	gallus gall
O13449	homo sapien
O81003	mus musculus
P20766	rattus norv
O05663	rattus norv
P20763	rattus norv
P01835	rattus norv
P15364	drosophila
P23533	mus musculus
O62813	rattus norv
P97685	rattus norv
P11627	mus musculus
O94856	homo sapien
P05538	homo sapien
P13765	homo sapien
O92477	mus musculus
P06731	homo sapien
P18467	pan troglod
O60469	homo sapien

ID	NAME	STANDARD	PRT	330 AA
107	135.5	5.6	105	1 LAC5_MOUSE
108	135.5	5.6	1266	1 NGCA_CHICK
109	135	5.6	837	1 NCW2_MOUSE
110	134.5	5.6	261	1 HB22_HUMAN
111	134.5	5.6	261	1 HB24_HUMAN
112	134.5	5.5	261	1 HB21_HUMAN
113	132.5	5.5	853	1 NCAL_BOVIN
114	132	5.5	231	1 HB21_CHICK
115	132	5.5	1302	1 NRG_BROWE
116	131.5	5.4	345	1 OPCW_BOVIN
117	131.5	5.4	345	1 OPCW_HUMAN
118	131	5.4	261	1 HB23_HUMAN
119	131	5.4	837	1 NCW2_HUMAN
120	131	5.4	858	1 NCAL_RAT
121	130.5	5.4	1234	1 NPHN_RAT
122	129	5.3	338	1 LAMP_CHICK
123	127.5	5.3	261	1 HB25_HUMAN
124	127.5	5.3	810	1 PLMN_ERIEU
125	127	5.3	725	1 NCA2_MOUSE

ALIGNMENTS

```

RESULT 1
GCL_HUMAN          STANDARD:      PRT:      330 AA.
AC   P01857;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Ig gamma-1 chain C region.
GN  IGHG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82274238; PubMed=6287432;
RA   Ellison J.W., Betson B.J., Hood L.E.;
RT   "The nucleotide sequence of a human immunoglobulin C gamma1 gene.";
RL   Nucleic Acids Res. 10:4071-4079(1982).
RN   [2]
RP   SEQUENCE OF 1-135 (MYELOMA PROTEIN EU).
RX   MEDLINE=71064024; PubMed=5489771;
RA   Cunningham B.A., Rutishauser U., Gall W.E., Gottlieb P.D.,
RA   Waxdal M.U., Edelman G.M.;
RT   "The covalent structure of a human gamma G-immunoglobulin. VII. Amino
RT   acid sequence of heavy-chain cyanogen bromide fragments H1-H4.";
RL   Biochemistry 9:3161-3170(1970).
RN   [3]
RP   SEQUENCE OF 136-329 (EU).
RX   MEDLINE=71064025; PubMed=5530842;
RA   Rutishauser U., Cunningham B.A., Bennett C., Konigsberg W.H.,
RA   Edelman G.M.;
RT   "The covalent structure of a human gamma G-immunoglobulin. 8. Amino
RT   acid sequence of heavy-chain cyanogen bromide fragments H5-H7.";
RL   Biochemistry 9:3171-3181(1970).
RN   [4]
RP   SEQUENCE (MYELOMA PROTEIN NIE).
RX   MEDLINE=77070269; PubMed=826475;
RA   Poncigl H., Hilschmann N.;
RT   "The rule of antibody structure. The primary structure of a
RT   monoclonal IgG1 immunoglobulin (myeloma protein Nie). III. The
RT   chymotryptic peptides of the H-chain, alignment of the tryptic
RT   peptide and discussion of the complete structure.";
RL   Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604(1976).
RN   [5]
RP   SEQUENCE (MYELOMA PROTEIN KOL) AND DISULFIDE BONDS.
RX   MEDLINE=83289131; PubMed=6884994;
RA   Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;
RT   "Three-dimensional structure determination of antibodies. Primary

```

```

RT   structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";
RT   Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).
RN   [6]
RP   DISULFIDE BONDS.
RX   MEDLINE=71064027; PubMed=4923144;
RA   Gall W.E., Edelman G.M.;
RT   "The covalent structure of a human gamma G-immunoglobulin. X.
RT   intrachain disulfide bonds.";
RL   Biochemistry 9:3188-3196(1970).
RN   [7]
RP   DISULFIDE BONDS.
RX   MEDLINE=77070267; PubMed=1002129;
RA   Dreker L., Schwertz J., Reichel W., Hilschmann N.;
RT   "Rule of antibody structure. The primary structure of a monoclonal
RT   IgG1 immunoglobulin (myeloma protein Nie). I: Purification and
RT   characterization of the protein, the L- and H-chains, the
RT   cyanogen bromide cleavage products, and the disulfide bridges.";
RL   Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX   MEDLINE=81208100; PubMed=7236608;
RA   Deisenhofer J.;
RT   "Crystallographic refinement and atomic models of a human Fe fragment
RT   and its complex with fragment B of protein A from Staphylococcus
RT   aureus at 2.9- and 2.8-A resolution.";
RL   Biochemistry 20:2361-2370(1981).
CC   -1- MISCELLANEOUS: Nie has the GIM(17) allotypic marker. 97-K, and the
CC   GIM(1) markers, 239-D and 241-L. KOL and EU sequences have the
CC   GIM(3) marker and the GIM (non-1) markers.
CC   -1- MISCELLANEOUS: Nie also differs in the amidation states of
CC   35, 116, 198, 269 and 272.
CC   -1- MISCELLANEOUS: EU also differs in the amidation states of residues
CC   155, 166, 177, 195, 198, 269, and 272 and in the order of residues
CC   268-272.
CC   -1- MISCELLANEOUS: KOL also differs in the amidation states of
CC   residues 198, 267 and 272.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; J00228; AAC82527.1; ALT_INIT.
DR   PIR; A93433; GHNU.
DR   PDB; 1FC1; 15-JUL-92.
DR   PDB; 1FC2; 15-JUL-92.
DR   PDB; 1AJ7; 12-NOV-97.
DR   PDB; 1D5B; 09-FEB-00.
DR   PDB; 1D5I; 09-FEB-00.
DR   PDB; 1D6V; 04-OCT-00.
DR   PDB; 1DN2; 17-MAY-00.
DR   PDB; 1E4K; 06-JUN-01.
DR   PDB; 1FCC; 20-JUL-95.
DR   PDB; 1H2H; 12-JUN-02.
DR   PDB; 1I7Z; 08-AUG-01.
DR   PDB; 1IIS; 16-MAY-01.
DR   PDB; 1IIX; 16-MAY-01.
DR   PDB; 1L6X; 10-APR-02.
DR   PDB; 2RCS; 12-NOV-97.
DR   Genew; HGNC:5525; IGHG1.
DR   MIM; 147100.
DR   GO; GO:0005624; C:membrane fraction; NAS.
DR   GO; GO:0003823; F:antigen binding; TAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003597; Ig cl.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF0047; Ig; 3.
DR   SMART; SM00407; IGc1; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.

```

DR PROSITE; PS00290; IG MHC; 2.
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 3D-structure.

FT	NON_TER	1	1
FT	DOMAIN	98	CH1.
FT	DOMAIN	99	HINGE.
FT	DOMAIN	111	CH2.
FT	DOMAIN	224	CH3.
FT	DISULFID	27	INTERCHAIN (WITH LIGHT CHAIN).
FT	DISULFID	103	INTERCHAIN (WITH HEAVY CHAIN).
FT	DISULFID	109	INTERCHAIN (WITH HEAVY CHAIN).
FT	DISULFID	112	INTERCHAIN (WITH HEAVY CHAIN).
FT	DISULFID	144	INTERCHAIN (WITH HEAVY CHAIN).
FT	DISULFID	250	INTERCHAIN (WITH HEAVY CHAIN).
FT	CARBOHYD	180	
FT	MOD_RES	330	
FT	VARIANT	97	
FT	VARIANT	239	
FT	VARIANT	241	
FT	STRAND	122	
FT	HELIX	130	
FT	TURN	136	
FT	STRAND	141	
FT	TURN	157	
FT	STRAND	163	
FT	TURN	165	
FT	STRAND	168	
FT	TURN	176	
FT	STRAND	180	
FT	TURN	182	
FT	HELIX	193	
FT	TURN	198	
FT	STRAND	202	
FT	TURN	209	
FT	STRAND	215	
FT	STRAND	227	
FT	TURN	230	
FT	HELIX	238	
FT	STRAND	245	
FT	STRAND	260	
FT	STRAND	270	
FT	STRAND	274	
FT	TURN	280	
FT	TURN	283	
FT	STRAND	287	
FT	HELIX	297	
FT	TURN	302	
FT	STRAND	305	
FT	TURN	313	
FT	TURN	316	
FT	TURN	325	
SO	SEQUENCE	330 AA; 36106 MW; 3770EBE106CFA3D CRC64;	

Query Match 50.8%; Score 1227.5; DB 1; Length 330;
 Best Local Similarity 74.5%; Pred. No. 1.1e-75;
 Matches 246; Conservative 7; Mismatches 22; Indels 55; Gaps 6;

Db NSTYRVAVSLVTLVHQLDWMNGKEVKCKVSKALPALEKTIISAKKQGPREFQYVTLPPSRD 239

Qy 180 NSTYRVAVSLVTLVHQLDWMNGKEVKCKVSKALPALEKTIISAKKQGPREFQYVTLPPSRD 239

Db 342 ELTKQVSLTCLVKGFFYPSDIAVEMWESNGQPENNYKTTTPVLDSGSPFLYSKLTVDKSR 401

Db 240 ELTKQVSLTCLVKGFFYPSDIAVEMWESNGQPENNYKTTTPVLDSGSPFLYSKLTVDKSR 299

Qy 402 WOQGNVFGSVWHEALNHNHYTKSLSPG 431

Db 300 WOQGNVFGSVWHEALNHNHYTKSLSPG 329

RESULT 2

GC3_HUMAN STANDARD; PRT; 290 AA.

AC P01860;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Ig gamma-3 chain C region (Heavy chain disease protein) (HDC).

GN IGHG3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE (DISEASE PROTEIN WIS).

RX MEDLINE=81021548; PubMed=6774747;

RA Frangione B., Rosenwasser E., Prelli F., Franklin E.C.;

RT "Primary structure of human gamma 3 immunoglobulin deletion mutant: gamma 3 heavy-chain disease protein WIS.";

RT Biochemistry 19:4304-4308(1980).

RN [2]

RP REVISIONS TO 12-97 (PROTEIN WIS).

RX MEDLINE=7118561; PubMed=402363;

RA Michaelson T.E., Frangione B., Franklin E.C.;

RT "Primary structure of the 'hinge' region of human IgG3. Probable quadruplication of a 15-amino acid residue basic unit.";

RT J. Biol. Chem. 252:863-869(1977).

RN [3]

RP REVISIONS TO 59-289 (PROTEIN WIS/DISEASE PROTEIN ZUC).

RX MEDLINE=77021516; PubMed=823945;

RA Wolfenstein-Todel C., Frangione B., Prelli F., Franklin E.C.;

RT "The amino acid sequence of 'heavy' chain disease' protein ZUC. Structure of the FC fragment of immunoglobulin G3.";

RT Biochem. Biophys. Res. Commun. 71:907-914(1976).

RN [4]

RP SEQUENCE FROM N.A. (DISEASE PROTEIN OMM).

RX MEDLINE=82247835; PubMed=6808505;

RA Alexander A., Steinmetz M., Barricault D., Frangione B., Franklin E.C., Hood L., Buxbaum J.N.;

RT "Gamma Heavy chain disease in man: cDNA sequence supports partial gene deletion model.";

RT Proc. Natl. Acad. Sci. U.S.A. 79:3260-3264(1982).

CC -1- SUNUNIT: Dimer linked by 12 disulfide bonds; it has an extra interchain disulfide bond at position 7 in addition to the 11 normally present in the hinge region.

CC -1- MISCELLANEOUS: The heavy chain disease protein WIS is shown.

CC -1- MISCELLANEOUS: The sequence of residues 42-76 was taken from the Ref. 2.

CC -1- MISCELLANEOUS: Disease protein WIS is lacking most of the V region and all of the CH1 region.

CC -1- MISCELLANEOUS: Disease protein ZUC lack most of the V region, all of the CH1 region, and part of the hinge compared with normal gamma-3 heavy chains.

CC -1- MISCELLANEOUS: Disease protein OMM may represent an allelic form or another gamma chain subclass.

CC -1- MISCELLANEOUS: The hinge region in gamma-3 chains is about four times as long as in other gamma chains and contains three identical 15-residue segments preceded by a similar 17-residue segment (12-28).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; J00231; AAA52805.1; ALT_SEQ.

DR HSSP; P01857; 1FC1.

DR Genew; HGNC:5527; IGHG3.

DR MIM; 147120; -

DR GO; GO:0005624; C:membrane fraction; NAS.

DR GO; GO:0003823; F:antigen binding; TAS.

DR GO; GO:0006955; P:immune response; NAS.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003597; Ig cl.

DR InterPro; IPR003006; Ig_MHC.

DR Pfam; PF00047; Ig; 2.

DR SMART; SM00407; IGc1; 1.

DR PROSITE; PS50835; IG LIKE; 2.

DR PROSITE; PS00290; IG_MHC; 1.

KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat;

Pyrolydione carboxylic acid.

FT DOMAIN 12 73 HINGE.

FT DOMAIN 74 183 CH2.

FT DOMAIN 184 289 CH3.

FT REPEAT 29 43

FT REPEAT 44 58

FT REPEAT 59 73

FT MOD_RES 1 1

FT CARBOHYD 6 6

FT DISULFID 7 7

FT DISULFID 24 24

FT DISULFID 27 27

FT DISULFID 33 33

FT DISULFID 39 39

FT DISULFID 42 42

FT DISULFID 48 48

FT DISULFID 54 54

FT DISULFID 57 57

FT DISULFID 63 63

FT DISULFID 69 69

FT DISULFID 72 72

FT DISULFID 140 140

FT MOD_RES 290 290

FT MOD_RES 126 127

FT VARIANT 134 134

FT VARIANT 139 139

FT VARIANT 182 182

FT VARIANT 227 227

FT VARIANT 227 227

FT VARIANT 227 227

FT VARIANT 279 279

FT VARIANT 279 279

FT SEQUENCE 290 AA; 3231 MM; E69CBG5705B2F46 CRC64;

Query Match 46.9%; Score 1133; DB 1; Length 290;

Best Local Similarity 88.2%; Pred. No. 2.1e-69;

Matches 209; Conservative 13; Mismatches 9; Indels 6; Gaps 2;

QY 201 PCP-APBPKSCDKHTC-----PELLGGSVFLFPFKPKDITLMTSRPEVTCVVVDVSHE 254

DB 53 PCPRCPBPKSCDTPPCPCPAPELLGGSVFLFPFKPKDITLMTSRPEVTCVVVDVSHE 112

QY 255 DPEVKNYVVGVEVHNAKTKREOVNSTYRVVSLVTLVLMODMNGEKYCKVSNKLP 314

DB 113 DPEVQFKMYVDGVVHNAKTKRPEOVNSTYRVVSLVTLVLMODMNGEKYCKVSNKLP 172

QY 315 APIEKTISKAKGPRPEVYTLTPSPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPEN 374

DB 173 APIEKTISKAKGPRPEVYTLTPSPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPEN 232

QY 375 NYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVPGSCVMEHALLNHYTQKSLSPG 431

DB 233 NYNTPPMLDSGDSFFLYSKLTVDKSRWQQGNVPGSCVMEHALLNHYTQKSLSPG 289

RESULT 3

ID GC2_HUMAN STANDARD; PRT; 326 AA.

AC P01859;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ig gamma-2 chain C region.

GN IGHG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxId=9606;

RN [1]

RP SEQUENCE OF 2-326 FROM N.A.

RX MEDLINE=82197621; PubMed=6804948;

RA Ellison J.W., Hood L.E.;

RT "Linkage and sequence homology of two human immunoglobulin gamma

heavy chain constant region genes";

RL Proc. Natl. Acad. Sci. U.S.A. 79:1984-1988(1982).

RN [2]

RP SEQUENCE OF 88-115 FROM N.A.

RX TISSUE=Fetal liver;

RA MEDLINE=83001943; PubMed=681139;

RT Takahashi N., Ueda S., Obata M., Nikaide T., Nakai S., Honjo T.;

RT "Structure of human immunoglobulin gamma genes: implications for

evolution of a gene family";

RL Cell 29:671-679(1982).

RN [3]

RP SEQUENCE OF 99-177 AND 310-326 FROM N.A.

RX TISSUE=Fetal liver;

RA MEDLINE=84235922; PubMed=6329676;

RT Krawinkel U., Rabblite T.H.;

RT "Comparison of the hinge-coding segments in human immunoglobulin gamma

heavy chain genes and the linkage of the gamma 2 and gamma 4 subclass

genes";

RL EMBO J. 1:403-407(1982).

RN [4]

RP SEQUENCE OF 1-325 (MYELOMA PROTEIN TIL).

RX MEDLINE=81007873; PubMed=6774012;

RA Wang A.-C., Tung E., Pudenberg H.H.;

RT "The primary structure of a human IgG2 heavy chain: genetic,

evolutionary, and functional implications";

RL J. Immunol. 125:1048-1054(1980).

RN [5]

RP SEQUENCE OF 1-85 AND 132-325 (MYELOMA PROTEIN ZIE).

RX MEDLINE=80001357; PubMed=113060;

RA Connell G.E., Parr D.M., Hofmann T.;

RT "The amino acid sequences of the three heavy chain constant region

domains of a human IgG2 myeloma protein";

RL Can. J. Biochem. 57:758-767(1979).

RN [6]

RP SEQUENCE OF 238-275 (ZIE).

RX MEDLINE=80114419; PubMed=118920;

RA Hofmann T., Parr D.M.;

RT "A note of the amino acid sequence of residues 361-391 of human

immunoglobulin gamma chains";

RL Mol. Immunol. 16:923-925(1979).

RN [7]

RP REVISIONS TO 25; 59; 60 AND 264-268 (ZIE).

RA Hofmann T., Parr D.M.;

RT Submitted (MAR-1980) to the PIR data bank.

RL [8]

RP SEQUENCE OF 1-121 (DOT).

```

RX MEDLINE=95255298; PubMed=7737190;
RA Scoppini M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;
RT "Characterization of the two unique human anti-Flavin monoclonal
RT immunoglobulins.",
RL Eur. J. Biochem. 228:886-893(1995).
RN (9)
RP DISULFIDE BONDS.
RX MEDLINE=72033500; PubMed=4940472;
RA Mlstein C., Frangione B.;
RT "Disulphide bridges of the heavy chain of human immunoglobulin G2.",
RL Biochem. J. 121:217-225(1971).
RN (10)
RP DISULFIDE BONDS.
RX MEDLINE=69064124; PubMed=5782707;
RA Frangione B., Mlstein C., Pink J.R.L.;
RT "Structural studies of immunoglobulin G.",
RL Nature 221:145-148(1969).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J060230; AAB59393.1; -.
DR PIR, A93906; G2HU.
DR HSSP, P01857; 1FC1.
DR GeneW, HGNC:5526; IGHG2.
DR MIM, 147110; -.
DR GO, GO:0005624; C:membrane fraction; NAS.
DR GO, GO:0003823; F:antigen binding; TAS.
DR GO, GO:0006955; P:immune response; NAS.
DR InterPro, IPR007110; Ig-1like.
DR InterPro, IPR003597; Ig_c1.
DR InterPro, IPR003006; Ig_MHC.
DR Pfam, PF00047; Ig_3.
DR SMART, SM00407; Igcl; 2.
DR PROSITE, PS50835; IG_LIKE; 3.
DR PROSITE, PS00290; IG_MHC; 2.
DR Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 1 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 229 CH2.
FT DOMAIN 220 326 CH3.
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 83
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 103 103 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT DISULFID 246 304
FT SITE 156 326
FT MOD_RES 326 326
FT VARIANT 60 60
FT CONFLICT 109 109 AT OR NEAR THE COMPLEMENT-BINDING SITE.
FT SEQUENCE 326 AA; 35884 MW; 8310878C6878CF9C CRC64; REMOVED POST-TRANSLATIONALLY (PROBABLE).
C -> S (IN REF. 3).
S -> A (IN MYELOMA PROTEINS TIL & ZIE).
/FTID=VAR 003889.
Query Match 46.8%; Score 1130.5; DB 1; Length 326;
Best Local Similarity 82.6%; Pred. No. 3.6e-69;
Matches 214; Conservative 12; Mismatches 12; Indels 21; Gaps 3;

```

```

QY 223 VLHODMLNGKEYCKCVSNKALPAPLEKTIISAKGQPREPOVYTLPPSRDELTKNOVSLTC 352
DB 187 VHQDMLNGKEYCKCVSNKNGKAPLEKTIISKGGPREPOVYTLPSBEMTKNOVSLTC 246
QY 353 LVKGFYPSDIAVWESNQGPNENYKTPPVLDSDSFFLYSKLTVDKSRMOQGNVFSGSV 412
DB 247 LVKGFYPSDIAVWESNQNQPNENYKTPPMLDSDSFFLYSKLTVDKSRMOQGNVFSGSV 306
QY 413 MHEALHNHYTKSKLSLSPG 431
DB 307 MHEALHNHYTKSKLSLSPG 325
RESULT 4
GC4_HUMAN
ID GC4_HUMAN STANDARD; PRT; 327 AA.
AC P01861;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-4 chain C region.
GN IGHG4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=83157104; PubMed=6299662;
RA Ellison J.W., Buxbaum J.N., Hood L.B.;
RT "Nucleotide sequence of a human immunoglobulin C gamma 4 gene.",
RL DNA 1:11-18(1981).
RN (2)
RP SEQUENCE OF 1-30 AND 81-326.
RX MEDLINE=70207560; PubMed=4192699;
RA Pink J.R.L., Buttery S.H., de Vries G.M., Mlstein C.;
RT "Human immunoglobulin subclases. Partial amino acid sequence of the
RT constant region of a gamma 4 chain.",
RL Biochem. J. 117:33-47(1970).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, K01316; AAB59394.1; ALT_INIT.
DR PIR, A90933; G4HU.
DR PDB, 1ADQ; 16-SEP-98.
DR GeneW, HGNC:5528; IGHG4.
DR MIM, 147130; -.
DR GO, GO:0005624; C:membrane fraction; NAS.
DR GO, GO:0003823; F:antigen binding; TAS.
DR GO, GO:0006955; P:immune response; NAS.
DR InterPro, IPR007110; Ig-1like.
DR InterPro, IPR003597; Ig_c1.
DR InterPro, IPR003006; Ig_MHC.
DR Pfam, PF00047; Ig_3.
DR SMART, SM00407; Igcl; 2.
DR PROSITE, PS50835; IG_LIKE; 3.
DR PROSITE, PS00290; IG_MHC; 2.
DR Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON_TER 1 1
FT DOMAIN 1 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 220 CH2.
FT DOMAIN 221 327 CH3.
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 83
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).

```

```

FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 141 201
FT DISULFID 247 305
SQ SEQUENCE 327 AA; 35940 MW; 3EBDB811EF208E7A CRC64;

Query Match 46.4%; Score 1119; DB 1; Length 327;
Best Local Similarity 83.7%; Pred. No. 2,1e-68;
Matches 215; Conservative 11; Mismatches 15; Indels 16; Gaps 4;

QY 181 TWTCV-----LQNKVPEFKIDVPCPAPRPKSCDTHCPCLLGGPSVFLPPPKD 234
DB 80 TYTCVNDHKPSNTKVDKRVESKYG-PPCP-----SC-----PAPEFLGGPSVFLPPPKD 129
QY 235 TLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEVHNKTKPRREQVSTYRVSVLTVL 294
DB 130 TLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEVHNKTKPRREQVSTYRVSVLTVL 189
QY 295 HQDWLNGKRYKCKVSNKALPAPIEKTIKSKAQPREPOVYTLPPSRDELTKNOVSLTCLV 354
DB 190 HQDWLNGKRYKCKVSNKALPAPIEKTIKSKAQPREPOVYTLPPSRDELTKNOVSLTCLV 249
QY 355 KGFYPSDIAVWESNGQPENNYKTPPYLDSDGSEFLYSKLTVDKSRNOQGNVRSQVYM 414
DB 250 KGFYPSDIAVWESNGQPENNYKTPPYLDSDGSEFLYSKLTVDKSRNOQGNVRSQVYM 309
QY 415 EALNHNHYOKSLSLSPG 431
DB 310 EALNHNHYOKSLSLSLG 326

RESULT 5
CD4_HUMAN STANDARD; PRT; 458 AA.
ID AC P01730;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85254948; PubMed=2990730;
RA Madden P.J., Littman D.R., Godfrey M., Madden D.E., Chess L.,
RA Axel R.;
RT "The isolation and nucleotide sequence of a cDNA encoding the T cell
RT surface protein T4: a new member of the immunoglobulin gene family.";
RL Cell 42:93-104(1985).
RN [2]
RP REVISION TO 26.
RX MEDLINE=89028665; PubMed=3263213;
RA Littman D.R., Madden P.J., Axel R.;
RT "Corrected CD4 sequence.";
RL Cell 55:541-541(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303695; PubMed=8723724;
RA Ancelet-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,
RA Malley T., Gibbs R.A.;
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase
RT genes at human chromosome 12p13.";
RL Genome Res. 6:314-326(1996).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANT TRP-265.
RX MEDLINE=91216786; PubMed=1708753;
RA Hodge T.W., Sasseo D.R., McDougal J.S.;
RT "Humans with OKT4-epitope deficiency have a single nucleotide base
RT change in the CD4 gene, resulting in substitution of TRP240 for
RT ARG240.";

```

```

RL Hum. Immunol. 30:99-104(1991).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Aleschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,
RA DiCicco L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Yoshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kelleman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalka U., Smalins D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [7]
RP SEQUENCE OF 26-394.
RX MEDLINE=90078232; PubMed=2592374;
RA Carr S.A., Hemling M.E., Folena-Wasserman G., Sweet R.W., Anumula K.,
RA Barr J.R., Huddleston M.J., Taylor P.;
RT "Protein and carbohydrate structural analysis of a recombinant
RT soluble CD4 receptor by mass spectrometry.";
RL J. Biol. Chem. 264:21286-21295(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 26-208.
RX MEDLINE=91061881; PubMed=1701030;
RA Wang J., Yan Y., Garrett T.P., Liu J., Rodgers D.W., Garlick R.L.,
RA Tarr G.E., Husain Y., Reinherz E.L., Harrison S.C.;
RT "Atomic structure of a fragment of human CD4 containing two
RT immunoglobulin-like domains.";
RL Nature 348:411-418(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 26-208.
RX MEDLINE=91061882; PubMed=2247146;
RA Ryu S.-E., Kwong P.D., Truneh A., Porter T.G., Athos J.,
RA Rothenberg M., Dai X., Xuong N.-H., Axel R., Sweet R.W.,
RA Hendrickson W.A.;
RT "Crystal structure of an HIV-binding recombinant fragment of human
RT CD4.";
RL Nature 348:419-426(1990).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (3.9 ANGSTROMS) OF 26-388.
RX MEDLINE=97311402; PubMed=9168119;
RA Wu H., Kwong P.D., Hendrickson W.A.;
RT "Dimeric association and segmental variability in the structure of
RT human CD4.";
RL Nature 387:527-530(1997).
RN [11]
RP PALMITOYLATION.
RX MEDLINE=92317088; PubMed=1618861;
RA Crise B., Rose J.K.;
RT "Identification of palmitoylation sites on CD4, the human
RT immunodeficiency virus receptor.";
RL J. Biol. Chem. 267:13593-13597(1992).

```

```

CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD4 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd4.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M12807; AAA35572.1; -
CC EMBL; U47924; AAB51309.1; -
CC EMBL; M31560; AAA16069.1; -
CC EMBL; BC025782; AAH25782.1; -
CC FIR; A90872; RWH074.
CC PDB; 1CDH; 30-APR-94.
CC PDB; 1CDI; 30-APR-94.
CC PDB; 3CD4; 31-OCT-93.
CC PDB; 1CDJ; 01-APR-97.
CC PDB; 1CDU; 01-APR-97.
CC PDB; 1CDY; 01-APR-97.
CC PDB; 1WER; 12-MAR-97.
CC PDB; 1WIO; 07-JUL-97.
CC PDB; 1WIP; 07-JUL-97.
CC PDB; 1WIO; 07-JUL-97.
CC PDB; 1G9M; 27-DEC-00.
CC PDB; 1G9N; 27-DEC-00.
CC PDB; 1G91; 19-AUG-98.
CC PDB; 1JL4; 19-SEP-01.
CC GLOSuiteDB; P01730; -
CC GeneW; HGNC;1678; CD4.
CC MIM; 166940; -
CC GO; GO:0042101; C:T-cell receptor complex; NAS.
CC GO; GO:0015026; F:coreceptor activity; NAS.
CC GO; GO:0015029; F:internalization receptor activity; TAS.
CC GO; GO:0042289; F:MHC class II protein binding; NAS.
CC GO; GO:0004889; F:transmembrane receptor activity; TAS.
CC GO; GO:0006955; P:immune response; NAS.
CC GO; GO:0009405; P:pathogenesis; TAS.
CC GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; NAS.
CC GO; GO:0030217; P:T-cell differentiation; NAS.
CC GO; GO:0045058; P:T-cell selection; NAS.
CC GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; NAS.
CC InterPro; IPR009713; CD4_TCS.
CC InterPro; IPR007110; Ig_Like.
CC Pfam; PF00047; Ig_2.
CC PRINTS; PR00692; CD4TCANTIGEN.
CC SMART; SM00406; IGV; 1.
CC PROSITE; PSS0835; IG_Like; 1.
CC Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
CC Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure;
CC Polymorphism.
CC SIGNAL 1 25
CC CHAIN 26 458
CC DOMAIN 26 396
CC TRANSMEM 397 418
CC DOMAIN 419 458
CC DOMAIN 26 125
CC DOMAIN 126 203
CC DOMAIN 204 317
CC DOMAIN 318 374
CC CARBOHYD 296
CC CARBOHYD 325

```

```

FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 470
FT LIPID 422 422
FT VARIANT 265 265
FT STRAND 27 32
FT TURN 33 34
FT STRAND 37 39
FT TURN 44 45
FT STRAND 51 55
FT TURN 56 57
Query Match 42.4%; Score 1023; DB 1; Length 458;
Best Local Similarity 99.5%; Pred. No. 9,5e-62;
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
BY SIMILARITY.
S-palmitoyl cysteine.
S-palmitoyl cysteine.
R -> W (in OKT4-negative populations).
/FTID=VAR_003906.
1 MKRGVPRHLVLVLLALPAATQGNKVVYLGKGGPTVELCTASQKKSIOFMKNSNOIK 60
1 MKRGVPRHLVLVLLALPAATQGNKVVYLGKGGPTVELCTASQKKSIOFMKNSNOIK 60
61 ILNGQSPFLTKGSPKLNDRASRRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQL 120
61 ILNGQSPFLTKGSPKLNDRASRRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQL 120
121 LVFGILTANSDTHLGGOSITLTLESPPGSSPSVQCRSPRGKNIQCGKTLVSQLELDSSG 180
121 LVFGILTANSDTHLGGOSITLTLESPPGSSPSVQCRSPRGKNIQCGKTLVSQLELDSSG 180
181 TWTCVTLOKQKVEFKIDIV 200
181 TWTCVTLOKQKVEFKIDIV 200
RESULT 6
CD4_PANTR STRAND; PRT; 458 AA.
ID CD4_PANTR STRAND; PRT; 458 AA.
AC P16004;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90182664; PubMed=2107024;
RA Cameron D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RT outside the virus binding site.";
RL Cell 60:747-754(1990).
RN [2]
RP SEQUENCE OF 26-424 FROM N.A.
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M3135; AAA5407.1; -
DR EMBL, X73323; CAA51749.1; -
DR PIR, B32722; RMC274.
DR HSSP, P01730; 1M10.
DR GO, GO:0042101; C1T-cell receptor complex; ISS.
DR GO, GO:0015026; F1coreceptor activity; ISS.
DR GO, GO:0042289; F1MHC class II protein binding; ISS.
DR GO, GO:0006955; P1immune response; ISS.
DR GO, GO:0045086; P1positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO, GO:0030217; P1T-cell differentiation; ISS.
DR GO, GO:0045058; P1T-cell selection; ISS.
DR GO, GO:0007169; P1transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPRO00973; CD4 TCAG.
DR InterPro; IPRO07110; Ig-like.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00406; CD4TCANTIGEN.
DR PROSITE; PS50835; IG_LIKE; 1.
DR Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
DR Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 42 42
FT CONFLICT 62 62
FT CONFLICT 191 191
SQ SEQUENCE 458 AA; 51057 MW; A7C3AC8A5257D3AD CRC64;

Query Match 41.4%; Score 999; DB 1; Length 458;
Beet Local Similarity 97.5%; Pred. No. 3,9e-60;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLTALPAATQGNKYVLGKGGTVELTCTAASKSIOGHMKNQIX 60
DB 1 MNRGVPFRHLVLTALPAATQGNKYVLGKGGTVELTCTAASKSIOGHMKNQIX 60
QY 61 ILNGSGFLTKGSPSLNDRADSRSLMDQGNPLIKLKIKEDSPTYCEVEDQKEEYQL 120
DB 61 ILNGSGFLTKGSPSLNDRADSRSLMDQGNPLIKLKIKEDSPTYCEVEDQKEEYQL 120
QY 121 LVFGLTANSDFHLLLOGSILTLTLESPPGSSPVQCRSPRGKNIQGGKTLASYOLELDSG 180
DB 121 LVFGLTANSDFHLLLOGSILTLTLESPPGSSPVQCRSPRGKNIQGGKTLASYOLELDSG 180
QY 181 TWTCTVLQNKVKVEFKIDIV 200
DB 181 TWTCTVLQNKVKVEFKIDIV 200

```

```

AC P03987;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DB Ig gamma-3 chain C region, membrane-bound form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85027161; PubMed=6092053;
RA Wells J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
RA Tucker P.W., Blatner F.R.;
RT "Structural analysis of the murine IgG3 constant region gene.";
RL EMO J. 3:2041-2046(1984).
RN [2]
RP SEQUENCE OF 328-398 FROM N.A.
RX MEDLINE=84041483; PubMed=6314258;
RA Komaromy M., Clayton L., Rogers J., Robertson S., Kettman J.,
RA Wall R.;
RT "The structure of the mouse immunoglobulin in gamma 3 membrane gene
RT segment.";
RL Nucleic Acids Res. 11:6775-6785(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00451; AAB59655.1; -
DR EMBL; V01526; CAA24767.1; ALT_SEQ.
DR PIR; A02156; GMSM.
DR HSSP; P01857; IFCL.
DR InterPro; IPRO07110; Ig-like.
DR InterPro; IPRO03597; IG_C1.
DR InterPro; IPRO03006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGCL; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
DR Transmembrane; Alternative splicing.
FT NON TER 1 1
FT DOMAIN 1 97
FT DOMAIN 98 113
FT DOMAIN 114 223
FT DOMAIN 224 327
FT TRANSMEM 346 362
FT DOMAIN 363 398
FT CONFLICT 333 333
FT CONFLICT 342 342
FT CONFLICT 388 388
SQ SEQUENCE 398 AA; 43929 MW; CF7F264B50A41B95 CRC64;

Query Match 39.6%; Score 955.5; DB 1; Length 398;
Beet Local Similarity 61.4%; Pred. No. 2,8e-57;
Matches 180; Conservative 40; Mismatches 58; Indels 15; Gaps 4;

QY 170 SVSOLELQDSGTW-----TCVLQNGKVEFKIDI--VPCPAPEPKSCDKHTCP--EL 219
DB 62 SLSSLVTPSSSTWPSQVVICNVAPASKTLIRIKRIKSTPPS-----SCPGNI 116
QY 220 LGGSVFLFPPKPKDITMISRTPEVTCVVVDVSHEDPEVKFNKYVDGVVNAATKPRE 279
DB 117 LGGSVFIFPPKPKDMLMISLTPEKVCVVVDVSEDDPDVAVSWVDKKEVHTAQTQREA 176
QY 280 QYNSTRVSVGLVTVLHDDMLNGEKYKKYKSNKALPAIETKISAKAQKQPEPVYTIPLPS 339
DB 177 QYNSTRVSVGLVTVLHDDMLNGEKYKKYKSNKALPAIETKISAKAQKQPEPVYTIPLPS 339

```



```

GN CDA.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymocytes;
RA Tatemura M., Yabe M., Yamada Y.K.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D63349; BAA09673.1; -.
DR HSSP: P01730; IMBR.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . . ; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 26 396 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 397 418 POTENTIAL.
FT DOMAIN 419 458 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 2.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 458 AA; 50872 MW; 9105479B5C56FF7 CRC64;
Query Match 38.0%; Score 916.5; DB 1; Length 458;
Beet Local Similarity 79.6%; Pred. No. 1.4e-54;
Matches 183; Conservative 14; Mismatches 28; Indels 5; Gaps 1;
Oy 1 MNRGVPFHLVLTALLPATGKNVVLGKGGDTVLTCTASQKSIQFHMKNNSQIK 60
Db 1 MNRGVPFHLVLTALLPATGKNVVLGKGGDTVLTCTASQKSIQFHMKNNSQIK 60
Oy 61 ILNGSFLTRKPSKLRADSRSLMDQGNFPIIKNLKTEDSDTYICEVEDQKEEVL 120

```

```

Db 61 ILNGSFLTRKPSKLRADSRSLMDQGNFPIIKNLKTEDSDTYICEVENKEEVL 120
Oy 121 LVFGLTANSTHLLQGSLTLTLESPPGSSPSVQCRSPRKRNIGGKTLTSSQLDSDG 180
Db 121 LVFGLTANSTHLLQGSLTLTLESPPGSSPSVQCRSPRKRNIGGKTLTSSQLDSDG 180
Oy 181 TWICTVQLQNKRYEFKIDIVPCPAPEPKSCDKHTHCPQLGSPVFLFPP 230
Db 181 TWICTVQLQNKRYEFKIDIVPCPAPEPKSCDKHTHCPQLGSPVFLFPP 225
RESULT 10
CD4_MACFU STANDARD; PRT; 458 AA.
AC P79184;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CDA.
OS Macaca fasciata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatemura M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D63348; BAA09672.1; -.
DR HSSP: P01730; IMBR.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . . ; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 26 396 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 397 418 POTENTIAL.
FT DOMAIN 419 458 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 2.

```

```

FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 458 AA; 50828 MW; 7683E7F08185535 CRC64;

Query Match 37.7%; Score 910; DB 1; Length 458;
Best Local Similarity 87.5%; Pred. No. 3.8e-54;
Matches 175; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 1 MNRGVPFRLLLVQLALPAPATQGNKVVLGKKGDTVELTCAQSKSIQFPMKNSNOIK 60
DB 1 MNRGVPFRLLLVQLALPAPATQGNKVVLGKKGDTVELTCAQSKSIQFPMKNSNOIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIDSPYICGEVDKEEYQL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIDSPYICGEVDKEEYQL 120
QY 121 LVFGLTANSDTHLQSGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLQSGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGLTANSDTHLQSGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLQSGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKVKVERKIDIV 200
DB 181 TWTCTVLQNKVKVERKIDIV 200
QY 181 TWTCTVLQNKVKVERKIDIV 200
DB 181 TWTCTVLQNKVKVERKIDIV 200

RESULT 11
GCAM_MOUSE STANDARD; PRT; 399 AA.
ID GCAM_MOUSE STANDARD; PRT; 399 AA.
AC P01865;
DT 21-JUL-1986 (Rel. 01, Created)
DR 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2A chain C region, membrane-bound form.
OS Mue musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82222190; PubMed=6283537;
RA Yamawaki-Kataoka Y., Nakai S., Miyata T., Honjo T.;
RT "Nucleotide sequences of gene segments encoding membrane domains of
RL immunoglobulin gamma chain.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627 (1982).
CC -1- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=Membrane-bound;
CC IsoId=P01865-1; Sequence=Displayed;
CC CC Name=Secreted;
CC CC IsoId=P01864-1; Sequence=External;
CC CC Note=Probably the major isoform;
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J00471; AAB59661.1; ALT_INTT.
DR PIR; A02154; G2MSAM.
DR PDB; 1KB5; 08-APR-98.
DR PDB; 1YEE; 15-OCT-97.
DR MGD; MGI:96443; Igh-1.

```

```

DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS00835; IG-LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane; Alternative splicing; 3D-structure; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 98 IG-LIKE 1.
FT DOMAIN 121 220 IG-LIKE 2.
FT DOMAIN 229 325 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 110 110 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 144 204
FT DISULFID 250 308
FT TRANSMEM 346 363 POTENTIAL.
FT CARBOHYD 180 399 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 399 AA; 44020 MW; 4C38138BAED3FF0 CRC64;

Query Match 37.6%; Score 906.5; DB 1; Length 399;
Best Local Similarity 60.5%; Pred. No. 5.5e-54;
Matches 173; Conservative 34; Mismatches 60; Indels 19; Gaps 5;

QY 179 SGTW-----TCTVQD--NQKVEFKID-----IVCPAPBPBSCDKTTCCELLGSPSVF 226
DB 72 SSTWPSQSIITCNVAPASTVVDKIEPRGPIKCP---PCKC-----PAPNLGGPSVF 124
QY 227 LPPPKQDTLMISRTPEVTCVVDVSHEDPEYKENVYDGYEVHNAKTRPBEQYNSTYR 286
DB 125 IFPPKIDKVLMSISPIVTCVVDVSEDDPVQISMFNANVEVHTAQOTRHREDYNSTLR 184
QY 287 VVSUTVTVHOMLNKCKKCVSNKALPAPLEKTSKAGQPREQVYTLPPSRDELTKN 346
DB 185 VVSALPIQHOMWSKEPKCKVNNKDLPAPIERTISKRGSVRAQVYVLPPEEEMTKK 244
QY 347 QVSLTCLVKGFPYSDIAVEMESNGCPENNYKTTTPVLDSDGSFPLYSKLTVDKSRMOQN 406
DB 245 QVTLTCMTVDMPEIIVYEWNNNGKTELANKTEBVLVSDGSYFYSGLRAREKKNWERN 304

RESULT 12
ID CD4_MACMU STANDARD; PRT; 458 AA.
AC P16003; Q29617;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90182664; PubMed=2107024;
RA Camerini D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RL outside the virus binding site.";
RN [2]

```

RP SEQUENCE FROM N.A.
 RC TISSUE=Thymocytes;
 RA Hashimoto O., Tatsumi M.;
 RT "Molecular cloning and expression of macaque CD4s";
 RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
 RN (3)
 RN SEQUENCE OF 28-424 FROM N.A.
 RC TISSUE=Blood;
 RA MEDLINE=93049640; PubMed=1425921;
 RT Fomsgaard A., Hirsch V.M., Johnson P.R.;
 RT "Cloning and sequences of primate CD4 molecules: diversity of the
 RT cellular receptor for simian immunodeficiency virus/human
 RT immunodeficiency virus";
 RL Eur. J. Immunol. 22:2973-2981(1992).
 [4]
 RP SEQUENCE OF 107-192 FROM N.A.
 RA MEDLINE=98320644; PubMed=9656488;
 RA Harris E.E., DiCoccia T.R.;
 RT "Nuclear gene trees and the phylogenetic relationships of the
 RT mangabeys (Primates: Papionini).";
 RL Mol. Biol. Evol. 15:892-900(1998).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lck (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.iesb-sib.ch/announce/>
 CC or send an email to license@iesb-sib.ch).
 CC -----
 CC EMBL: M31134; AAA36838.1; -;
 DR EMBL: D63347; BA09671.1; -;
 DR EMBL: X73326; CAA51752.1; -;
 DR EMBL: AF057385; AAC25129.1; -;
 DR HSPB; P01730; 1MBR.
 DR GO; GO:0042101; C-T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.
 DR GO; GO:0042289; F:MHC class II protein binding; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO; GO:0030217; P:T-cell differentiation; ISS.
 DR GO; GO:0045058; P:T-cell selection; ISS.
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
 DR InterPro: IPR000973; CD4 TCAG.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003596; IG_v.
 DR Pfam: PF00047; Ig_2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG LIKE; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW immune response; Repeat; Signal; Lipoprotein; Palmitate.
 FT SIGNAL 1 25
 FT CHAIN 26 458
 FT DOMAIN 26 396
 FT TRANSMEM 397 418
 FT DOMAIN 419 458
 FT DOMAIN 26 125
 FT DOMAIN 126 203
 FT DOMAIN 204 317
 FT DOMAIN 318 374
 FT CARBOHYD 296 296
 FT CARBOHYD 325 325
 FT DISULFID 41 109
 FT DISULFID 155 184
 FT DISULFID 328 370
 FT LIPID 419 419
 S-palmitoyl cysteine (By similarity).

FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
 FT CONFLICT 42 42 N -> T (IN REF. 1).
 FT CONFLICT 62 62 L -> S (IN REF. 3).
 FT CONFLICT 67 67 I -> S (IN REF. 2).
 FT CONFLICT 169 169 I -> L (IN REF. 2).
 FT CONFLICT 191 191 K -> N (IN REF. 3).
 FT CONFLICT 248 248 S -> P (IN REF. 2).
 FT CONFLICT 265 265 R -> Q (IN REF. 3).
 FT CONFLICT 349 349 A -> T (IN REF. 2).
 SQ SEQUENCE 458 AA; 50884 MW; 8BB80339FAFEC808 CRC64;
 Query Match 37.4%; Score 904; DB 1; Length 458;
 Best Local Similarity 87.0%; Pred. No. 9,7e-54;
 Matches 174; Conservative 13; Mismatches 13; Indels 0; Gaps 0;
 QY 1 MNRGVRPHLLVYQALLPAATQGNKVLTGKQDPTVELTCTASOKRSIQFHMNSNOIK 60
 DB 1 MNRGIPRHLILVQLLPAVTOGKKVAGKGDVETLTGNSQKNTQFHMNSNOIK 60
 QY 61 ILNGQSFLTKGSPKLNDRADSRSLMDQGNPFLIKNLKTEDSDTYICEVEDQKEVQL 120
 DB 61 ILGIQGLFKGSPKLSDRADSRSLMDQCFMIIKNIKEDSDTYICEVENKKEVEL 120
 QY 121 LVFGLTANSDTHLLQGSILTLTESPPGSSPVQCRSPKKNIQGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLQGSILTLTESPPGSSPVQCRSPKKNIQGKRTISVPLERDQSG 180
 QY 181 TWTCVTYQNKVYEFKIDIV 200
 DB 181 TWTCVTYSQDKTVEFKIDIV 200
 RESULT 13
 ID CD4_MACNE STANDARD; PRT; 458 AA.
 AC Q08340; P79196;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
 DE T4/Leu-3).
 GN CD4.
 OS Macaca nemestrina (Pig-tailed macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 CX NCBI_TaxID=9545;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hashimoto O., Tatsumi M.;
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 28-424 FROM N.A.
 RC TISSUE=Blood;
 RA MEDLINE=93049640; PubMed=1425921;
 RT Fomsgaard A., Hirsch V.M., Johnson P.R.;
 RT "Cloning and sequences of primate CD4 molecules: diversity of the
 RT cellular receptor for simian immunodeficiency virus/human
 RT immunodeficiency virus";
 RL Eur. J. Immunol. 22:2973-2981(1992).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lck (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.iesb-sib.ch/announce/>

```

CC or send an email to license@sb-sib.ch).
CC EMBL: D63346; BAA09670.1; -.
CC EMBL: X73325; CAA51751.1; -.
DR HSSP: P01730; IWER.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; P:coreceptor activity; ISS.
DR GO: GO:0042289; P:coreceptor binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:003217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4 TCAG.
DR InterPro: IPR007110; Ig-Like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PS50835; IG_Like; 1.
KM Immunoglobulin domain, Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT CARBOHYD 318 374
FT CARBOHYD 42 42
FT CARBOHYD 296 326
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 57 57
FT CONFLICT 91 91
FT CONFLICT 105 105
FT CONFLICT 113 113
FT CONFLICT 302 302
FT CONFLICT 349 349
SQ SEQUENCE 458 AA; 50905 MW; 751A9BA2CB83BE16 CRC64;

Query Match 37.4%; Score 903; DB 1; Length 458;
Best Local Similarity 87.0%; Pred. No. 1,1e-53;
Matches 174; Conservative 14; Mismatches 12; Indels 0; Gaps 0;

QY 1 NMRGPVFRHLVLVQLALPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHWKNSNOIK 60
DB 1 NMRGPVFRHLVLVQLALPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHWKNSNOIK 60
QY 61 ILGNQSPFLTKGPKSKINDRARSRLMDQGNPLLIQNKIKIDSDPYICEVDQKEEYQL 120
DB 61 ILGIQSPFLTKGPKSKINDRARSRLMDQGNPLLIQNKIKIDSDPYICEVDQKEEYQL 120
QY 121 LVFGLTANSDTHLQOQSILTLTLESPGSSPSVQCSPPGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLQOQSILTLTLESPGSSPSVQCSPPGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLYNQKQKVEFKIDIV 200
DB 181 TWTCTVLYNQKQKVEFKIDIV 200

```

```

DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84030930; Pubmed=6313520;
RA Bernstein K.E., Alexander C.B., Mage R.G.;
RT "Nucleotide sequence of a rabbit IgG heavy chain from the recombinant
RT F-1 haplotype."
RL Immunogenetics 16:387-397(1983).
RN [2]
RP SEQUENCE OF 1-128.
RX MEDLINE=76135469; Pubmed=1243651;
RA Pratt D.M., Mole L.E.;
RT "Sequence studies on the constant region of the Fd sections of rabbit
RT immunoglobulin G of different allotype.",
RL Biochem. J. 151:337-349(1975).
RN [3]
RP SEQUENCE OF 88-266 FROM N.A.
RX MEDLINE=83299917; Pubmed=6193512;
RA Martens C.L., Moore K.W., Steinmetz M., Hood L., Knight K.L.;
RT "Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma
RT heavy chain and identification of two genomic C gamma genes."
RL Proc. Natl. Acad. Sci. U.S.A. 79:6018-6022(1982).
RN [4]
RP SEQUENCE OF 132-161.
RX MEDLINE=70110015; Pubmed=5461106;
RA Fruchter R.G., Jackson S.A., Mole L.E., Porter R.R.;
RT "Sequence studies of the Fd section of the heavy chain of rabbit
RT immunoglobulin G."
RL Biochem. J. 116:249-259(1970).
RN [5]
RP SEQUENCE OF 129-131 AND 155-322.
RA Hill R.L., Lebovitz H.E., Fellows R.E. Jr., Delaney R.;
RL (In) Killander J. (eds.);
RL Gamma globulins, Nobel symp. 3, pp.109-127, Almqvist and Wiksell,
RL Stockholm (1967).
CC -I- MISCELLANEOUS: Ref.1 sequence has the D12 allelic marker,
CC 104-Thr, and the E14 marker, 185-Thr. Ref.3 has the D11 and E15
CC markers and Ref.5 the E15 marker.
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC
CC EMBL: M16426; AAA31289.1; -.
DR PIR: A81749; GHRB.
DR HSSP: P01857; IFCT.
DR InterPro: IPR007110; Ig-1-like.
DR InterPro: IPR003597; Ig_c1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_3.
DR SMART: SM00407; IGV; 2.
DR PROSITE: PS50835; IG_Like; 3.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT DOMAIN 1 1
FT DOMAIN 6 96
FT DOMAIN 114 213
FT DOMAIN 222 318
FT VARIANT 104 104
FT VARIANT 185 185
FT CONFLICT 48 48
FT CONFLICT 71 71

```


DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Griwet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OK NCBI_TaxID=9534;
RN [1]
RN SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RT "Molecular cloning and expression of african green monkey CD4.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 28-424 FROM N.A.
RP TISSUE=Blood;
RC MEDLINE=93049640; PubMed=1425921;
RX Fomsgaard A., Hirsch V.M., Johnson P.R.;
RA "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [3]
RN SEQUENCE OF 28-424 FROM N.A.
RP TISSUE=Periphereal blood;
RC MEDLINE=98017879; PubMed=9379478;
RX Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RN [4]
RN SEQUENCE OF 107-192 FROM N.A.
RP MEDLINE=98320644; PubMed=9656488;
RX Harris E.E., Diocetell T.R.;
RA "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (Primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).

CC
EMBL, D86589, BAA13132.1, -;
DR EMBL, X73322, CAA51748.1, -;
DR EMBL, AF001226, AAB60873.1, -;
DR EMBL, AF001228, AAB60875.1, -;
DR EMBL, AF057380, AAC25124.1, -;
DR HSSP, P01730, 1W1O.
DR GO, GO:0042101, C:T-cell receptor complex, ISS.
DR GO, GO:0015026, F:coreceptor activity, ISS.
DR GO, GO:0042289, F:MHC class II protein binding, ISS.
DR GO, GO:0006955, P:immune response, ISS.
DR GO, GO:0045086, P:positive regulation of interleukin-2 biosyn. . ., ISS.
DR GO, GO:0030217, P:T-cell differentiation, ISS.
DR GO, GO:0045058, P:T-cell selection, ISS.
DR GO, GO:0007169, P:transmembrane receptor protein tyrosine kin. . ., ISS.
DR InterPro, IPR000973, CD4_TCSA.
DR InterPro, IPR007110, Ig-like.
DR InterPro, IPR003596, Ig_v.
DR Pfam, PF000477, Ig, 2.
DR PRINTS, PR00692, CD4TCANTIGEN.
DR SMART, SM00406, IGV, 1.

```

RX MEDLINE=82115295; PubMed=6799207;
RA Rogers J., Choi E., Souza L., Carter C., Word C.J., Kuehl M.,
RA Eigenberg D., Wall R.;
RT "Gene segments encoding transmembrane carboxyl termini of
RT immunoglobulin gamma chains.";
RL Cell 26:19-27(1981).
CC
CC -1- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Membrane-bound;
CC IsoId=PI01867-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=PI01866-1; Sequence=External;
CC Note=May be the major isoform;
CC -1- PTM: DISULFIDE BONDS BY HOMOLOG WITH OTHER IG GAMMA CHAINS.
CC -1- MISCELLANEOUS: The sequence of residues 1-335 is assumed to be
CC identical with the corresponding region of the secreted form.
CC -1- MISCELLANEOUS: The a allele sequence is shown.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J00462; AAB59659.1; ALT_INIT.
CC PIR: C02154; G2MSBM.
CC PDB: 1C1C; 11-MAR-03.
CC MGD: MGI:96445; Igh-3.
CC DR InterPro: IPR007110; IG-1like.
CC DR InterPro: IPR003597; IG_C1.
CC DR InterPro: IPR003006; IG_MHC.
CC Pfam: PF00047; Ig; 3.
CC SMART: SM00407; IgC1; 2.
CC DR PROSITE: PS50835; IG_LIKE; 3.
CC DR PROSITE: PS00290; IG_MHC; 1.
CC KW Immunoglobulin domain; Immunoglobulin C region; Transmembrane;
CC Alternative splicing; 3D-structure; Repeat.
CC
CC FT NON TER 1
CC FT DOMAIN 6 98 IG-LIKE 1.
CC FT DOMAIN 127 226 IG-LIKE 2.
CC FT DOMAIN 235 331 IG-LIKE 3.
CC FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
CC FT DISULFID 27 82 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 118 118 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 150 210
CC FT DISULFID 256 314
CC FT TRANSMEM 352 369
CC FT DOMAIN 370 405 POTENTIAL.
CC FT SEQUENCE 405 AA; 44330 MW; 89B3CF0A9B6D9FA CRC64;
SQ
Query Match 36.4%; Score 877.5; DB 1; Length 405;
Best Local Similarity 51.7%; Pred. No. 5e-52;
Matches 178; Conservative 52; Mismatches 75; Indels 39; Gaps 9;
OY 133 LIQG--OSLTITLSPGSS-----PSVQCRSPRGKNIQGGKTLVSQLELDQSGTW-- 182
DB LVKGFPESSVTVVWNGSLSSSVHTFPAL-----LQSG-LVTWSSSVTVSSSTPPS 77
OY 183 ---TCTVQ--NQKKVEFKID-----IVCPAPPEPKSCDKHTPC--PELLGGSVLEF 228
DB QTVVTSVAHPASSTTVDKLPSGISTINPCP-----PCKECHCCPAPNLEGGSVTF 132
OY 229 PPKPKDTLMISRTPEVTCVVDVSHDEPEVKNVYDVEVNAKTKPREQYNSTYREV 288
DB 117 LGGSVLEFPKPKDMLMISLTPEKTCVVDVSDPDDVQISWVNNVHTAQTQVHREDYNSTIRV 192
OY 289 SVLTIVLHODMLNGEKYKCKVSNKALPAIEKTISSAKQAPREPOVYTLPPSRDELTKNQV 348
```

```

DB 193 STLPIDHQDMWSSGKEFKCKNNKNDLPSPIERTITISKIGLVRAPOVYILPPPAEQLSRKDV 252
OY 349 SLTCLVGVGFPSPIDAVWESNGQPENNYKTTPTPLDSDGSFPLYSKLTIVKSRMQQNVF 408
DB 253 SLTCLVGVGFPSPIDAVWESNGQPENNYKTTPTPLDSDGSFPLYSKLTIVKSRMQQNVF 312
OY 409 SCGVMEBALNNHYTQKSLSPGLQDDETCAEADQDGLDGLMTT 452
DB 313 SCVVRHGLKNVYLKKTISRSPLDLDLDCAEAKDGLDGLMTT 356
RESULT 18
ID3_MOUSE STANDARD; PRT; 329 AA.
AC P22436;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig gamma-3 chain C region, secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=65027161; PubMed=6092053;
RA Wells J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
RA Tucker P.W., Blattner F.R.;
RT "Structural analysis of the murine IgG3 constant region gene.";
RL EMBL J. 3:2041-2046(1984).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J00451; -. NOT_ANNOTATED_CDS.
CC PIR: B02156; G3MSC.
CC DR HSSP: P01857; IFC1.
CC DR InterPro: IPR007110; IG-1like.
CC DR InterPro: IPR003597; IG_C1.
CC DR InterPro: IPR003006; IG_MHC.
CC Pfam: PF00047; Ig; 3.
CC SMART: SM00407; IgC1; 2.
CC DR PROSITE: PS50835; IG_LIKE; 3.
CC DR PROSITE: PS00290; IG_MHC; 1.
CC KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
CC Transmembrane; Alternative splicing.
CC
CC FT NON TER 1
CC FT DOMAIN 1 97 CH1.
CC FT DOMAIN 98 113 HINGE.
CC FT DOMAIN 114 223 CH2.
CC FT DOMAIN 224 327 CH3.
CC FT SEQUENCE 329 AA; 36228 MW; F45827174182BAD6 CRC64;
SQ
Query Match 35.5%; Score 857.5; DB 1; Length 329;
Best Local Similarity 59.6%; Pred. No. 8.6e-51;
Matches 162; Conservative 38; Mismatches 57; Indels 15; Gaps 4;
OY 170 SVSOLQLDQSGTW-----TCTVQNGKVEFKID---VCPAPPEPKSCDKHTPC--EL 219
DB 62 SLSSLVTVPSSTWPSQVIVCVNAHPASKTLIKRIERIKFSPFPS-----SCPQNT 116
OY 220 LGGSVLEFPKPKDMLMISRTPEVTCVVDVSHDEPEVKNVYDVEVNAKTKPREE 279
DB 117 LGGSVLEFPKPKDMLMISLTPEKTCVVDVSDPDDVQISWVNNVHTAQTQVHREDYNSTIRV 176
OY 280 QYNSTYVSVLTIVLHODMLNGEKYKCKVSNKALPAIEKTISSAKQAPREPOVYTLPPS 339
```

```
Db 177 QYNSTFRVVSALPIQHODMMRGKFKCVNNKALPAPIERTISKPKRAQTPOVYTIPPP 236
QY 340 RELTKNOVSLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDK 399
Db 237 REOMSKKKVSLTCLVTFNPFSEAI SVEMENGELEDDYKTPPLDSDGYFLYSKLTVD 296
QY 400 SRMOQGNVFCSCVMHEALHNHYTOKSLSPG 431
Db 297 DSMLOGEIFTCVVMHEALHNHYTOKSLSPG 328

RESULT 19
GCB_RAT ID GCB_RAT STANDARD; PRT; 333 AA.
AC P20761;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG gamma-2b chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OK NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Bruggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
CC -1 SIMILARITY: Contains 3 immunoglobulin-like domains.
DR PIR; PS0018; PS0018.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGcl; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 96 IG-LIKE 1.
FT DOMAIN 124 223 IG-LIKE 2.
FT DOMAIN 232 328 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 80 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 147 207
FT DISULFID 253 311
SQ SEQUENCE 333 AA; 36497 MW; 55F8B64D8D460A6 CRC64;

Query Match 34.5%; Score 833; DB 1; Length 333;
Best Local Similarity 58.6%; Pred. No. 3.9e-49;
Matches 157; Conservative 38; Mismatches 53; Indels 20; Gaps 5;

179 SGTW-----TCTVLQ--NOKVVEFKIDI-----VPCPAPEPKSCDKHTTC--PELLGSP 223
Db 70 SSTWSSQVITCNVAVPASSTKVDKVERNNGIGHKCP-----TQPTCHKCVPPELLGSP 124
QY 224 SYFLPPPKPDITLMISRTPEVTCVVYDVSHEDPEVKFNMYVDGVEYHNAKTPREBEQYNS 283
Db 125 SVFLPPPKPDITLLISQNAKVCVVVDVSEEPDVQVFWVNNVEVHTAQTPREBEQYNS 184
QY 284 TVRVSVLTVLHODMNLGKCYKCVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDEL 343
Db 165 TRRVSVLALPIQHODMMRGKFKCVNNKALPAPIEKTISKPKRAQTPOVYTIPPP 244
QY 344 TNQVSVLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQ 403
Db 245 TQQTAVSLTCLTSGFLPNDIGVEMTNSGHIETKNTPEPMDSDGSPFMYSKLTVNRSRMD 304
```

```
QY 404 QGNVFCSCVMHEALHNHYTOKSLSPG 431
Db 305 SRAPFCVVMHEGLHNHYTOKSLSPG 332

RESULT 20
GCL_RAT ID GCL_RAT STANDARD; PRT; 326 AA.
AC P20759;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG gamma-1 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OK NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Bruggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
DR PIR; PS0017; PS0017.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGcl; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 112 HINGE.
FT DOMAIN 113 219 CH2.
FT DOMAIN 220 326 CH3.
FT DISULFID 27 82
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT DISULFID 246 304
FT CARBOHYD 176 176
SQ SEQUENCE 326 AA; 35946 MW; 013BAB45EF49B9DA CRC64;

Query Match 34.1%; Score 823.5; DB 1; Length 326;
Best Local Similarity 50.0%; Pred. No. 1.7e-48;
Matches 163; Conservative 53; Mismatches 55; Indels 55; Gaps 9;

133 LLOG---OSLTLTLESPGSS-----PSVQCRSPRGKNIQGGKTLVSQLELDSDGTW-- 182
Db 28 LVKGYFPEPVYVITNMGSLSSGVHTFPV-----LQGLGLYLTLSVTV-PESTIPS 77
QY 183 ---TCTVLQ--NOKVVEFKIDI VPCPAPEPKSCDKHTTCPELLG-----BSV 225
Db 78 QVTVCNVAVPASSTKVDKDI-----VPRNC-----GDDCKPCICTGSEVSSV 119
QY 226 FLPPPKPDITLMISRTPEVTCVVYDVSHEDPEVKFNMYVDGVEYHNAKTPREBEQYNS 285
Db 120 FLPPPKPDVLTITLTPKTCVVVDISQDDPEVHSMFVDVEVHTAQTPREBEQFNSTF 179
QY 286 RVSVSVLTVLHODMNLGKCYKCVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTK 345
Db 180 RSVSVLPIHODMNLGKTRFCKTISAAPPSIETKISCPBEQTPVPHYVTSPTKEBNTQ 239
QY 346 NQVSVLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQ 405
Db 240 NQVSVLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQ 299
```

Qy 406 NVFSCSVHHEALHNHYTKSLSLSPG 431
 Db 300 NTFCTSVHHEGLHNHHTKSLSHSPG 325

RESULT 21

GCC_RAT STANDARD; PRT; 329 AA.
 ID AC P20762;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 36, Last annotation update)
 DE Ig gamma-2c chain C region.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88166903; PubMed=3127222;
 RA Bruggemann M., Delmastro-Galfre P., Waldmann H., Calabi F.;
 RT "Sequence of a rat immunoglobulin gamma 2c heavy chain constant
 region cDNA: extensive homology to mouse gamma 3.",
 RL Eur. J. Immunol. 18:317-319(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X07189; CA30169.1; -.
 DR PIR, S00847; S00847.
 DR HSSP, P01842; 7EAB.
 DR InterPro; IPR007110; IG-1like.
 DR InterPro; IPR003597; IG_C1.
 DR InterPro; IPR003006; IG_MHC.
 DR Pfam, PF00047; Ig_2.
 DR SMART, SM00407; IGL1; 2.
 DR PROSITE; PS50835; IG_LIKE; 3.
 DR PROSITE; PS00290; IG_MHC; 1.
 KW Immunoglobulin domain; Immunoglobulin C region.
 FT NON TER 1 1
 FT DOMAIN 1 97 CH1.
 FT DOMAIN 98 113 HINGE.
 FT DOMAIN 114 222 CH2.
 FT DOMAIN 223 329 CH3.
 FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
 FT DISULFID 27 82
 FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 113 113 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 143 203
 FT DISULFID 249 307
 SQ SEQUENCE 329 AA; 36571 MW; 5FCD7B7933850773 CRC64;
 Query Match 34.1%; Score 823; DB 1; Length 329;
 Best Local Similarity 58.5%; Pred. No. 1.8e-48;
 Matches 151; Conservative 45; Mismatches 56; Indels 6; Gaps 2;
 Qy 179 SGTW-----TCTVQONQKVFETIDIVCPAPEPKSCDKHTCEELGGPSEVFLPFPKPK 233
 Db 72 SSTWSSQVTCVSVAHPATKSNLKIETP-RRPKRPPTDSCDDNLRGPRVFTFPKPK 130
 Qy 234 DTLWISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTV 293
 Db 131 DILMTLTPLPKTCVVVDVSEEPQVGQFMFVNDRVFAQTPQPHBEQNGFRVRSSTLHI 190
 Qy 294 LHWOMLNKPKYKCKSNKALPAPKEFKITISKAKGQPREPQVYTLTPSRBELTKNOVSLTCL 353
 Db 191 QHODMSGKERKCKVNNKDLPSPIETKISKPRGAKARTQVYTTIPPRQMSKNKVSILTCM 250

Qy 354 VKGYPEDIAVWESNCGPENNNTKTPPVLDSDGSEFLYKSLTVDKSRMOQGNVFCSCVM 413
 Db 251 VTSEYPAISIVEMERNCELEQDYKNTLPVLDSDSESYLYKSLVSDTDSWNRGDIYTCSSV 310

Qy 414 HEALHNHYTKSLSLSPG 431
 Db 311 HEALHNHYTKSLSLSPG 328

RESULT 22

GCI_MOUSE STANDARD; PRT; 324 AA.
 ID AC P01868;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ig gamma-1 chain C region secreted form.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80045036; PubMed=115593;
 RA Honjo T., Obata M., Yamawaki-Kataoka Y., Kataoka T., Kawakami T.,
 RA Takahashi N., Mano Y.;
 RT "Cloning and complete nucleotide sequence of mouse immunoglobulin
 RT gamma 1 chain gene";
 RL Cell 18:559-568(1979).
 RN [2]
 RP SEQUENCE OF 76-324 FROM N.A. (MYELOMA PROTEIN MOPC 31C).
 RX MEDLINE=8020559; PubMed=6769752;
 RA Obata Y., Yamawaki-Kataoka Y., Takahashi N., Kataoka T., Shimizu A.,
 RA Mano Y., Seidman J.G., Peterlin B.M., Leder P., Honjo T.;
 RT "Immunoglobulin gamma 1 heavy chain gene: structural gene sequences
 RT cloned in a bacterial plasmid";
 RL Gene 9:87-97(1980).
 RN [3]
 RP SEQUENCE OF 70-322 FROM N.A. (MYELOMA PROTEIN MOPC 21).
 RX MEDLINE=80012837; PubMed=113776;
 RA Rogers J., Clarke P., Salsber W.;
 RT "Sequence analysis of cloned cDNA encoding part of an immunoglobulin
 RT heavy chain";
 RL Nucleic Acids Res. 6:3305-3321(1979).
 RN [4]
 RP SEQUENCE (MYELOMA PROTEIN MOPC 21).
 RX MEDLINE=78242288; PubMed=98524;
 RA Adetugbo K.;
 RT "Evolution of immunoglobulin subclasses. Primary structure of a
 RT murine myeloma gamma1 chain";
 RL J. Biol. Chem. 253:6068-6075(1978).
 RN [5]
 RP DISULFIDE BONDS (MOPC 21).
 RX MEDLINE=73008889; PubMed=5073237;
 RA Svaszt J., Milstein C.;
 RT "The disulphide bridges of a mouse immunoglobulin G1 protein";
 RL Biochem. J. 126:837-850(1972).
 RL SUBCELLULAR LOCATION: Secreted.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Secreted;
 CC IsoId=P01868-1; Sequence=Displayed;
 CC Note=May be the major isoform;
 CC Name=Membrane-bound;
 CC IsoId=P01869-1; Sequence=External;
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL; V00793; CAA24172.1; -
DR EMBL; V00793; CAA24173.1; -
DR EMBL; V00793; CAA24174.1; -
DR EMBL; V00793; CAA24175.1; -
DR EMBL; V00795; CAA24176.1; -
DR PIR; A02159; G1MS.
DR GlycosuiteDB; P01868; -.
DR MGI; 96446; Igh-4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing.
FT NON_TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 110 HINGE.
FT DOMAIN 111 217 CH2.
FT DOMAIN 218 324 CH3.
FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .).
FT DISULFID 244 302 /FTid=CAR_000055.
FT MOD_RES 324 324 REMOVED POST-TRANSLATIONALLY.
FT CONFLICT 276 276 N -> D (IN REF. 3).
FT CONFLICT 278 278 N -> D (IN REF. 3).
SQ SEQUENCE 324 AA; 35704 MW; A338812F3D1F2C93 CRC64;

Query Match 33 9%; Score 818.5; DB 1; Length 324;
Beet Local Similarity 51.8%; Pred. No. 3,6e-48;
Matches 156; Conservative 50; Mismatches 60; Indels 35; Gaps 5;

QY 133 LKGGSLTLTLESPGSSPVOCRSRPGKNIQGGKTLVSQLELDPGWTCTVLQ--NQ 190
DB 56 VLQSLVLTSSSVTPSSP-----RSEIVTCNVAAHPASS 90

QY 191 KKEVERKIDIVPCPAPRPSKCDKTHCTPELLGSPVFLPPPKKDTLMISRTPEVTCVVD 250
DB 91 TVYDKKIVPRDCGC-KPCIC-----TVPEV---SSVFIFPPKPKDVLITLTPKTCVVVD 142

QY 251 VSHEDPEVKYKVVGVGVENNAKTKPREQVNSTVRVSVLVLVLMODMNGEKYCKVSN 310
DB 143 ISKDDPEVQSFVVDVEVHTAQIQREQFNSTRVSSELTIMQDMWNGEPEFCRVNS 202

QY 311 KALPAPIEKTIISKAGQPREPOVYTLPRSRDELTKQVSLTCLVGVGYPSDIAVEMESNG 370
DB 203 AAFPAPIEKTIISKTKRPAKAPVYITIPPKEDMAKDVKSLTCLMIDFFEDITVERQWNG 262

QY 371 QPENNYKTPPVLDSDGSEFLYSKLTVDKSRMQQGNVSCVMEALNNHTYQKSLSLSP 430
DB 263 QPAENYKNTQPIIMNTGSYFYVSKLVNQSNWEAGNTFTCSYLHGLNHNHTKSLSHSP 322

QY 431 G 431
DB 323 G 323

RESULT 23
GCAA MOUSE STANDARD; PRT; 330 AA.
AC F01863;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE Ig gamma-2A chain C region, A allele.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=81076554; PubMed=6777755;
RA Sikorav J.-L., Auffray C., Rougeon F.;
RT "Structure of the constant and 3' untranslated regions of the murine
RL Balb/c gamma 2a heavy chain messenger RNA."
RN Nucleic Acids Res. 8:3143-3155(1980).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=81198976; PubMed=6262729;
RA Yamawaki-Kataoka Y., Miyata T., Honjo T.;
RT "The complete nucleotide sequence of mouse immunoglobulin gamma 2a gene
RT and evolution of heavy chain genes: further evidence for intervening
RT sequence-mediated domain transfer."
RN Nucleic Acids Res. 9:1365-1381(1981).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=81223894; PubMed=6787604;
RA Olio R., Auffray C., Morchamps C., Rougeon F.;
RT "Comparison of mouse immunoglobulin gamma 2a and gamma 2b chain genes
RT suggests that exons can be exchanged between genes in a multigenic
RL family."
RL Proc. Natl. Acad. Sci. U.S.A. 78:2442-2446(1981).
[4]
RP MYELOMA PROTEIN MOPC 173.
RX MEDLINE=74175517; PubMed=4831970;
RA Bourgois A., Fougereau M., Rocca-Serra J.;
RT "Determination of the primary structure of a mouse IgG2a
RT immunoglobulin: amino-acid sequence of the Fc fragment. Implications
RT for the evolution of immunoglobulin structure and function."
RL Eur. J. Biochem. 43:423-435(1974).
[5]
RP DISULFIDE BONDS.
RX MEDLINE=73056887; PubMed=4565406;
RA de Preval C., Fougereau M.;
RT "Determination of the primary structure of a mouse gamma G2a
RT immunoglobulin. Identification of the disulfide bridges."
RL Eur. J. Biochem. 30:452-462(1972).
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DB EMBL; V00798; CAA24178.1; -.
DR PIR; A02152; G2MSA.
DR PDB; 1E4W; 12-JUL-01.
DR PDB; 1E4X; 12-JUL-01.
DR PDB; 1MN0; 06-MAY-99.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure; Repeat.
FT NON_TER 1 1
FT DOMAIN 1 98 IG-LIKE 1.
FT DOMAIN 121 220 IG-LIKE 2.
FT DOMAIN 229 325 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).

```

```
FT DISULFID 110 110 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 144 204
FT DISULFID 250 308
FT MOD_RES 330 330
SQ SEQUENCE 330 AA; 36389 MW; B84361C5445A6864 CRC64;

Query Match 33.4%; Score 805.5; DB 1; Length 330;
Best Local Similarity 58.5%; Pred. No. 2.7e-47;
Matches 155; Conservative 33; Mismatches 56; Indels 19; Gaps 5;

QY 179 SGTW-----TCTVLQ--NQKVEFKID-----IVPCAPAPKSCDKHTTCPELLGSPSVF 226
D 72 SSTWPSQGITCNVAHPASTVKDKIEPRGPTIKPCF---PCKC---PANNLLGGSPVF 124
QY 227 LFPKPKDGLMISTPEVTCVVDVSHDEPKFKNWYDGVFNNAKTKPREEOYNSTYR 286
D 125 IFPKPKDGLMISTPEVTCVVDVSHDEPKFKNWYDGVFNNAKTKPREEOYNSTYR 184
QY 287 VVSVLTVLHOMLNGKCKYKSNKALPAPIEKTSKAGQPREPOVYTLPPSRDELTKN 346
D 185 VVSALPIHODMNGKCKYKSNKALPAPIEKTSKAGQPREPOVYTLPPSRDELTKN 244
QY 347 QVSLTCLVKGFPSPDIAMWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGN 406
D 245 QVTLTCWTVTDMPEDIVYEMTNGKTELNYKTEPVLDSDGSFFLYSKLTVDKSRMOQGN 304
QY 407 VFSCSVNHEALHNHYTKSKLSLSPG 431
D 305 SYSQSVNHEALHNHYTKSKLSLSPG 329

RESULT 24
GCAB_MOUSE STANDARD; PRT; 335 AA.
ID GCAB_MOUSE STANDARD; PRT; 335 AA.
AC P01862;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-2A chain C region secreted form (B allele).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=82037861; PubMed=6170065;
RA Schreier P.H., Bothwell A.L.M., Mueller-Hill B., Baltimore D.;
RT "Multiple differences between the nucleic acid sequences of the
RT IgG2a and IgG2b alleles of the mouse.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:4495-4499(1981).
RN [2]
SEQUENCE.
RX MEDLINE=82037777; PubMed=6794027;
RA Dognin M.J., Lauwereys M., Strosberg A.D.;
RT "Multiple amino acid substitutions between murine gamma 2a heavy
RT chain Fc regions of Ig1a and Ig1b alleotypic forms.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:4031-4035(1981).
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Secreted;
CC IsoId=P01864-1; Sequence=Displayed;
CC Name=Membrane-bound;
CC Note=Probably the major isoform;
CC IsoId=P01865-1; Sequence=External;
CC -1- MISCELLANEOUS: The sequence differs from that of the a allele,
CC from BALB/c mice, at 15% of the positions.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J00479; -; NOT_ANNOTATED_CDS.
DR EMBL; A02153; G2MSAB.
DR PDB; 1BOG; 23-MAR-89.
DR PDB; 1HH6; 26-JAN-01.
DR PDB; 1HH9; 24-JUL-03.
DR PDB; 1HI6; 08-FEB-01.
DR InterPro; IPR007110; IG-Like.
DR InterPro; IPR003597; IG cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PSS0835; IG_LIKE; 3.
DR PROSITE; PSS0290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Alternative splicing;
KM 3d-structure; Repeat.
FT NON_TER 1
FT DOMAIN 6 98 IG-LIKE 1.
FT DOMAIN 126 225 IG-LIKE 2.
FT DOMAIN 234 330 IG-LIKE 3.
SQ SEQUENCE 335 AA; 36596 MW; FA3382792CB813C6 CRC64;

Query Match 33.2%; Score 801; DB 1; Length 335;
Best Local Similarity 50.2%; Pred. No. 5.ee-47;
Matches 166; Conservative 52; Mismatches 77; Indels 36; Gaps 9;

QY 124 GLTPNSDT--HLQG---QSLTTLLESPGSS-----PSVQCRSPRGKNIQGGKTLVSQ 173
D 17 GTTGSSVTLGCLVKGFPPEVTLTMNSGSLSSGVTHTPAL-----LQSG-LYTLSS 66
QY 174 LEIDSGTW-----TCTV-----LONQKVEFKIDIV--PCAPAPKSCDKHTTCPELL 220
D 67 SVTSTWTSQGITCNVAHPASTVKDKIEPRVPTIQPCF---PHQRPVPCAPDPL 123
QY 221 GGSVFLFPKPKDGLMISTPEVTCVVDVSHDEPKFKNWYDGVFNNAKTKPREEQ 280
D 124 GGSVFLFPKPKDGLMISTPEVTCVVDVSHDEPKFKNWYDGVFNNAKTKPREEQ 183
QY 281 YNSTYRVSVLTVLHOMLNGKCKYKSNKALPAPIEKTSKAGQPREPOVYTLPPSR 340
D 184 YNSTLRVVSALPIHODMNGKCKYKSNKALPAPIEKTSKAGQPREPOVYTLPPSR 243
QY 341 DELTKQVSLTCLVKGFPSPDIAMWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKS 400
D 244 EEMTKKEFSLTCMTGFLPAEIAVDWTSNGRTQNYKTATVLDSDGSFFLYSKLTVOKS 303
QY 401 RMOQGNVFSCSVNHEALHNHYTKSKLSLSPG 431
D 304 TWERGSLFACSVNHEALHNHYTKSKLSLSPG 334

RESULT 25
GCA_RAT STANDARD; PRT; 322 AA.
ID GCA_RAT STANDARD; PRT; 322 AA.
AC P20760;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2A chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Brueggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
```

```

RL Gene 74:473-482(1988).
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M3804; AAA41376.1; ALT_INIT.
CC PIR: P50019; P50019.
CC HSSP: P01842; 7EAB.
CC DR InterPro: IPR007110; Ig_Like.
CC DR InterPro: IPR003597; Ig_C1.
CC DR InterPro: IPR003006; Ig_MHC.
CC DR Pfam: PF00047; Ig_2.
CC DR SMART: SM00407; IgC1; 2.
CC DR PROSITE: PS00835; IG_Like; 3.
CC DR PROSITE: PS00290; IG_MHC; 1.
CC KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
CC FT NON_TER 1 1
CC FT DOMAIN 6 98 IG-Like 1.
CC FT DOMAIN 115 232 IG-Like 2.
CC FT DOMAIN 221 317 IG-Like 3.
CC FT DISULFID 27 82
CC FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 105 105 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 136 196
CC FT DISULFID 242 300
CC FT CARBOHYD 172 172
CC SQ SEQUENCE 322 AA; 35186 MW; E8EA136A9DE01EDB CRC64;
Query Match 32.8%; Score 792.5; DB 1; Length 322;
Best Local Similarity 51.0%; Pred. No. 2e-46;
Matches 160; Conservative 47; Mismatches 72; Indels 35; Gaps 8;
QY 133 LLOG---QSLTLTLESPSS-----PSYQCRSPRGKNIQGGKTLTSVQLELQDSGTW-- 182
DB 28 LVGVPEPEVTVWNGALSSGVHTPPV-----LQSLGYLTSSVTV-PSSTWGS 77
QY 183 ---TCTVQL--NOKTKEFKIDIVPCAPAPBKSCDKHTPELLGGSVLVFPKPKDTLM 237
DB 78 QAVTCVVAHPASTKYDKKIVPREC---NFCGTGSEV-----SSVEIPFPKTKDVL 127
QY 238 ISRTPEVTCVWVDSHEDPEVKFNWYVVDGVEVHNATKPREQVNSTYVWVSVLTVLHOD 297
DB 128 ITLTPKVTGVVDISQNDPEVRSMVIDVEVHTAQTAAPEKQSNSTLRSVSELPVHRD 187
QY 298 WLNKGEYKCKVSKALPAPIETKISKAKQOPREPOVYTLPPSRDELTKQVSLTCLVKGF 357
DB 188 WLNKGTFFCKVNSGAPAPIEKSIKPEGRPRGPVYTAAPREEMTQSOVSTICWVGF 247
QY 358 YPSDIAVESNGQEPENNTKTPPVLDSDGSFFLYSKLTVDKSRQGVNFGCVWHEAL 417
DB 248 YPDPDIETWKRMNGQEPENYKNTPTDSDGYFLYKLVKKETQWQGGFTFCVLAHEGL 307
QY 418 HHHYTKSLSPG 431
DB 308 HHHHTKSLSHSPG 321

```

```

GN CD4.
OS Cercopithecus torquatus atys (Red-crowned mangabey) (Sooty mangabey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=95311;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA MEDLINE=93049640; PubMed=1425921;
RX Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus";
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -I- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -I- SUBUNIT: Associates with p56-ick (By similarity).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73328; CAAS1754.1; -.
CC DR EMBL: X73327; CAAS1753.1; -.
CC DR HSSP: P01730; 1MTQ.
CC DR GO: GO:0042101; C:T-cell receptor complex; ISS.
CC DR GO: GO:0015026; F:coreceptor activity; ISS.
CC DR GO: GO:0042289; F:MHC class II protein binding; ISS.
CC DR GO: GO:0006955; P:immune response; ISS.
CC DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
CC DR GO: GO:0030217; P:T-cell differentiation; ISS.
CC DR GO: GO:0045058; P:T-cell selection; ISS.
CC DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
CC DR InterPro: IPR000973; CD4 TCAG.
CC DR InterPro: IPR007110; Ig-Like.
CC DR InterPro: IPR003596; Ig_V.
CC DR Pfam: PF00047; Ig_2.
CC DR PRINTS: PR00692; CD4TCANTIGEN.
CC DR SMART: SM00406; IGV; 1.
CC DR PROSITE: PS00835; IG_Like; 1.
CC KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
CC Immune response; Repeat; Lipoprotein; Palmitate.
CC FT NON_TER 1 1
CC FT DOMAIN <1 369 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 370 391 POTENTIAL.
CC FT DOMAIN 392 >397 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN <1 98 IG-Like V-TYPE.
CC FT DOMAIN 99 176 IG-Like C2-TYPE 1.
CC FT DOMAIN 177 290 IG-Like C2-TYPE 2.
CC FT DOMAIN 291 347 IG-Like C2-TYPE 3.
CC FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT DISULFID 14 82 BY SIMILARITY.
CC FT DISULFID 128 157 BY SIMILARITY.
CC FT DISULFID 301 343 BY SIMILARITY.
CC FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
CC FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
CC FT VARIANT 20 20 MISSING.
CC FT VARIANT 43 43 T -> I.
CC FT VARIANT 46 46 N -> D.
CC FT VARIANT 86 86 V -> L.
CC FT VARIANT 96 96 F -> L.
CC FT VARIANT 173 173 V -> M.
CC FT VARIANT 316 316 R -> K.
CC FT NON_TER 397 397

```

```

SQ SEQUENCE 397 AA; 43926 MW; 8660B636D2B38A7 CRC64;
Query Match 32.5%; Score 784; DB 1; Length 397;
Best Local Similarity 87.3%; Pred. No. 9.7e-46;
Matches 151; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 28 VVLGKGDVTELTCTASQKSIQFHWKNSNQIKILGNQGSFLTQKPSKLNDRADSRSLW 87
DB 1 VVLGKGDVTELTCTASQKSIQFHWKNSNQIKILGNQGSFLTQKPSKLNDRADSRSLW 60
QY 68 DQGNFPLIINKLKIEDSDTYICEVEDQKEVOLLVFGLTANSPTHLQGGSLTTLTLESP 147
DB 61 DQGFSEMIINKLKIEDSETTYICEVENKKEVEVLVFGLTANSPTHLQGGSLTTLTLESP 120
QY 148 GSSPSVOCRSRPRGNKIOGGKTLVSQLELDQSGTCTVYLOQKVEFKIDIV 200
DB 121 GSSPSVOCRSRPRGNKIOGGKTLVSQLELDQSGTCTVYLOQKVEFKIDIV 173

RESULT 27
CD4_ERYPA STANDARD; PRT; 397 AA.
AC 008339;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
DE (Fragment).
GN CD4.
OS Erythrocebus patas (Red guenon) (Hussar).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Erythrocebus.
OC NCBI_TaxID=9538;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus."
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -!- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -!- SUBUNIT: Associates with p56-lck (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X73324; CAA51750.1; -.
DR HSBP; P01730; IWO.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4 TAG.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; DATCANTIGEN.

```

```

DR SMART; SMO0406; IGV; 1.
DR PROSITE; PS0835; IG LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW immune response; Repeat; Lipoprotein; Palmitate.
FT NON_TER 1
FT DOMAIN 1 369 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 370 391 POTENTIAL.
FT DOMAIN 392 397 CYTOPLASMIC (POTENTIAL).
FT DOMAIN >397 IG-LIKE V-TYPE.
FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 291 347 IG-LIKE C2-TYPE 3.
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT 298 298 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT NON_TER 397
SQ SEQUENCE 397 AA; 44081 MW; 67887397A6B7EA4F CRC64;
Query Match 32.4%; Score 783; DB 1; Length 397;
Best Local Similarity 86.7%; Pred. No. 1.1e-45;
Matches 150; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 28 VVLGKGDVTELTCTASQKSIQFHWKNSNQIKILGNQGSFLTQKPSKLNDRADSRSLW 87
DB 1 VVLGKGDVTELTCTASQKSIQFHWKNSNQIKILGNQGSFLTQKPSKLNDRADSRSLW 60
QY 88 DQGNFPLIINKLKIEDSDTYICEVEDQKEVOLLVFGLTANSPTHLQGGSLTTLTLESP 147
DB 61 DQGFSEMIINKLKIEDSETTYICEVENKKEVEVLVFGLTANSPTHLQGGSLTTLTLESP 120
QY 148 GSSPSVOCRSRPRGNKIOGGKTLVSQLELDQSGTCTVYLOQKVEFKIDIV 200
DB 121 GSSPSVOCRSRPRGNKIOGGKTLVSQLELDQSGTCTVYLOQKVEFKIDIV 173

RESULT 28
GCB_MOUSE STANDARD; PRT; 336 AA.
ID GCB_MOUSE
AC P01866;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-2B chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ALIELE A).
RX MEDLINE=80120716; PubMed=676534;
RA Yamawaki-Kataoka Y., Kataoka T., Takahashi N., Oba M., Honjo T.;
RT "Complete nucleotide sequence of immunoglobulin gamma2b chain gene
RT cloned from newborn mouse DNA."
RL Nature 283:786-789(1980).
RN [2]
RP SEQUENCE FROM N.A. (MPC 11).
RX MEDLINE=80081501; PubMed=117548;
RA Tucker P.W., Marcu K.B., Slichtom J.L., Blattner F.R.;
RT "Structure of the constant and 3' untranslated regions of the murine
RT gamma 2b heavy chain messenger RNA."
RL Science 206:1299-1303(1979).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80081502; PubMed=117549;
RA Tucker P.W., Marcu K.B., Newell N., Richards J., Blattner F.R.;
RT "Sequence of the cloned gene for the constant region of murine gamma
RT 2b immunoglobulin heavy chain."
RL Science 206:1303-1306(1979).
RN [4]

```

RP SEQUENCE FROM N.A. (ALLELE B).
RX MEDLINE=82173203; PubMed=6803173;
RA Ollio R., Rougeon F.;
RT "Mouse immunoglobulin allotypes: post-duplication divergence of gamma
2a and gamma 2b chain genes.";
RL Nature 296:761-763(1982).
RN (5)
RP CARBOHYDRATE-LINKAGE SITE THR-105.
RX MEDLINE=94216359; PubMed=7512967;
RA Klm H., Yamaguchi Y., Maeda K., Matsunaga C., Yamamoto K.,
Irimura T., Takahashi N., Kato K., Arata Y.;
RT "O-glycosylation in hinge region of mouse immunoglobulin G2b.";
RL J. Biol. Chem. 269:12345-12350(1994).
CC -1- SUBCELLULAR LOCATION: Secreted (potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Secreted;
CC IsoId=PI01866-1; Sequence=Displayed;
CC Note=May be the major isoform;
CC Name=Membrane-bound;
CC IsoId=PI01867-1; Sequence=External;
CC -1- PTM: O-LINKED GLYCAN CONSISTS OF GAL-GALNAc DISACCHARIDE WHICH IS
MODIFIED WITH 2 SULFIC ACID RESIDUES.
CC -1- PTM: DISULFIDE BONDS BY HOMOLGY WITH OTHER IG GAMMA CHAINS.
CC -1- MISCELLANEOUS: The a allele sequence is shown.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DR PIR, S25057; G2MS11.
DR HSP; P01842; 7FAB.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Repeat.
FT DOMAIN 1 1
FT NON TER 1 1
FT DOMAIN 6 96
FT DOMAIN 127 226
FT DOMAIN 235 331
FT DISULFID 15 15
FT DISULFID 27 82
FT DISULFID 109 109
FT DISULFID 112 112
FT DISULFID 115 115
FT DISULFID 118 118
FT DISULFID 150 210
FT DISULFID 256 314
FT DISULFID 314 314
FT CARBOHYD 105 105
FT MOD_RES 336 336
FT VARIANT 163 163
FT VARIANT 194 194
FT VARIANT 300 300
FT VARIANT 301 301
FT VARIANT 301 301
FT CONFLICT 25 25
FT CONFLICT 36 36
FT CONFLICT 239 239
SQ SEQUENCE 336 AA; 3658 MW; 7D879662607C356E CRC64;

O-LINKED (GALNAc...)
REMOVED POST-TRANSLATIONALLY (PROBABLE).
Q -> R (IN ALLELE B).
T -> A (IN ALLELE B).
N -> D (IN ALLELE B).
M -> I (IN ALLELE B).
L -> S (IN REF. 2 AND 3).
S -> P (IN REF. 2 AND 3).
I -> T (IN REF. 2 AND 3).

Query Match 32.4%; Score 781.5; DB 1; Length 336;
Best Local Similarity 49.8%; Pred. No. 1,2e-45;
Matches 161; Conservative 50; Mismatches 73; Indels 39; Gaps 9;

DB 133 PENIDVLMISLTPKVTGVVDVSDDDPVOQSWVNVNVEHTAQTHREDYNTIRV 192
QY 289 SVLTVLHODWLNKGEYKCKVKSKALPAPIEKTISKAKQPREPVYTLPPSDELTKQV 348
DB 193 STLPIDQHDMNSGKEFKCKVNNKDLPSPIERTISKIKGLVRAQVYILPPPAEQLSRKV 252
QY 349 SLTCLVKGFPEPDIVNEPNSQPENNYKTPPVLDSCGSFELYEKLTVDSKRMQGNV 408
DB 253 SLTCLVGNPNPDIVNEPNSQPENNYKTPPVLDSCGSFELYEKLTVDSKRMQGNV 312
QY 409 SCNVNHEALNNHYTKSLSPG 431
DB 313 SCNVNHEALNNHYTKSLSPG 335

RESULT 29
CD4 SAISC STANDARD; PRT; 457 AA.
ID CD4 SAISC
AC 029037;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX NCBI_Taxid=9521;
RN (1)
RP SEQUENCE FROM N.A.
RA Tatum M., Hashimoto O.;
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D86588; BA1131.1; -.
DR HSP; P01730; IWER.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 457
FT DOMAIN 26 395
FT TRANSMEM 396 417
FT DOMAIN 418 457
CYTOPLASMIC (POTENTIAL).
CYTOSOL (POTENTIAL).

```

FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 202 IG-LIKE C2-TYPE 1.
FT DOMAIN 203 316 IG-LIKE C2-TYPE 2.
FT DOMAIN 317 373 IG-LIKE C2-TYPE 3.
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 327 369 BY SIMILARITY.
FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 457 AA; 50871 MW; 57EDB634005A015 CRC64;

Query Match
Best Local Similarity 71.5%; Pred. No. 1.6e-42;
Matches 143; Conservative 24; Mismatches 32; Indels 1; Gaps 1;

QY 1 MNRGVPFRHLIVLQALIPAAATQGNKVYLGGKGDVTELTCTASQKSIQFHMKNNOIK 60
DB 1 NMGGIPFRHLIVLQALIPAAATQGNKVYLGGKGDVTELTCTASQKSIQFHMKNNOIK 60
QY 61 ILNGSGFLTKGSPKLNDRADSRRLMDQGNFPLIKNLKTEDSTTYCEVEDQKEEYOL 120
DB 61 ILGVGNVFTVTRGQSKLTDRIDSKSSWDGSPFLIKARLEDSSTYICEVSKKEEVEL 120
QY 121 LVFGLTANSDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 QVFGLTANPDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 121 QVFGLTANPDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 QVFGLTANPDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWCTVTLNQKKVEFKIDIV 200
DB 181 TWCTVFOHLELV-PEINIV 199
QY 181 TWCTVTLNQKKVEFKIDIV 200
DB 181 TWCTVFOHLELV-PEINIV 199

RESULT 30
CD4_CANFA STANDARD; PRT; 463 AA.
ID CD4_CANFA STANDARD; PRT; 463 AA.
AC P33705;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell) surface antigen
GN CD4.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OC NCBI_TaxId=9615;
OK NCBI_TaxId=9615;
RN SEQUENCE OF 13-463 FROM N.A.
RP STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=93192324; PubMed=7916632;
RA Mide K.F., Conner G.E., Minz D.H., Alejandre R.;
RT "Primary structure of the canine CD4 antigen.";
RL Biochim. Biophys. Acta 1172:315-318(1993).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
RT alpha antigens.";
RL Tissue Antigens 43:184-188(1994).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in macrophages and a subset of
CC T lymphocytes.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L06130; AAB02295.1; -.
DR EMBL; X68565; -; NOT_ANNOTATED_CDS.
DR HSPSP; P01730; IMBR.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; P: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0006955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: Transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00692; CD4TCANTIGEN.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 24
FT CHAIN 25 463
FT DOMAIN 25 401 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 402 423 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 424 463 IG-LIKE V-TYPE.
FT DOMAIN 125 211 IG-LIKE C2-TYPE 1.
FT DOMAIN 212 321 IG-LIKE C2-TYPE 2.
FT DOMAIN 322 378 IG-LIKE C2-TYPE 3.
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 332 374 BY SIMILARITY.
FT LIPID 424 424 S-palmitoyl cysteine (By similarity).
FT LIPID 427 427 S-palmitoyl cysteine (By similarity).
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 463 AA; 51639 MW; 95805170CB44A833 CRC64;

Query Match
Best Local Similarity 34.7%; Score 610.5; DB 1; Length 463;
Matches 162; Conservative 62; Mismatches 130; Indels 113; Gaps 15;

QY 1 MNRGVPFRHLIVLQALIPAAATQGNKVYLGGKGDVTELTCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPFRHLIVLQALIPAAATQGNKVYLGGKGDVTELTCTASQKSIQFHMKNNOIK 60
QY 61 ILNGSGFLTKGSPKLNDRADSRRLMDQGNFPLIKNLKTEDSTTYCEVEDQKEEYOL 120
DB 61 ILNGSGFLTKGSPKLNDRADSRRLMDQGNFPLIKNLKTEDSTTYCEVEDQKEEYOL 120
QY 121 LVFGLTANSDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 171
DB 121 LVFGLTANSDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 171
QY 121 LVFGLTANSDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 171
DB 121 LVFGLTANSDTHLQGGSLTTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 171
QY 172 SQLELQDSGVTCTVTONQKKVEFKIDIVPCAPAPKSCDKTTCPLLGSPVFLPPK 231
DB 172 SQLELQDSGVTCTVTONQKKVEFKIDIVPCAPAPKSCDKTTCPLLGSPVFLPPK 231
QY 180 SWPELQGGTWTCTIISQSKTVEINIVLAF--QKVNTPYARE--GQVVEFSP-- 232
DB 180 SWPELQGGTWTCTIISQSKTVEINIVLAF--QKVNTPYARE--GQVVEFSP-- 232
QY 232 PKDTLMSRTPVTCVVVDVSHEDPEV--KFNWYVDGVEVHANKTKREBQYNSTYRV 289
DB 232 PKDTLMSRTPVTCVVVDVSHEDPEV--KFNWYVDGVEVHANKTKREBQYNSTYRV 289
QY 233 -----LSFEDENLVGELRWQAQAS-----SS 254
DB 233 -----LSFEDENLVGELRWQAQAS-----SS 254

```

```

QY 290 VLTVLHQMVLNGKEYKCKVSNKALPAPLEKTIISKAGQPRE--PQVYTLPPSRDELTKNQ 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 LL-----WISFLERKLSMKENAPL-----KLQMKESIPLFTLPOULSRASG 301
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 348 VSLTCLVKGFPSPDIAVEMESNGQPNKYTPPVLSDGSFPLYSKLTVDKSRMOQGNV 407
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 302 ILTLNLAKGTLGYDEV-----NLVVMRANSSQNNL 330
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 408 FSGSVNHEALHNHYTKSLSLSPGLQDLETCAADGELDLMTTPP 454
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 -TCEVLGVP-----TSPBLTL--LNLKEQAQVSKR-QOKLWVWVDP 367
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 31
CD4_RABIT STANDARD; PRT; 459 AA.
AC P46530;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
   T4/Leu-3).
GN CD4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92390370; PubMed=1518821;
RA Hague B.F., Sawasdi Kosol S., Brown T.J., Lee K., Recker D.P.,
RT Kindt T.J.,
RT "CD4 and its role in infection of rabbit cell lines by human
   immunodeficiency virus type 1."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7963-7967 (1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
   receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation-
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC
CC EMBL: M92840; AAA1198.1; -.
DR PIR: A46254; A46254.
DR HSSP: P01730; 1MBR.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro: IPR000973; CD4_TCRG.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_Like_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 459 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 26 396 EXTRACELLULAR (POTENTIAL).

```

```

FT TRANSMEM 397 419 POTENTIAL.
FT DOMAIN 420 459 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 129 IG-LIKE V-TYPE.
FT DOMAIN 130 208 IG-LIKE C2-TYPE 1.
FT DOMAIN 209 318 IG-LIKE C2-TYPE 2.
FT DOMAIN 319 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 41 113 POTENTIAL.
FT DISULFID 329 370 POTENTIAL.
FT LIPID 420 420 S-palmitoyl cysteine (By similarity).
FT LIPID 423 423 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 459 AA; 50886 MW; B323311CBDA0013D CRC64;

Query Match 24.9%; Score 601.5; DB 1; Length 459;
Best Local Similarity 41.9%; Pred. No. 2,2e-33;
Matches 153; Conservative 47; Mismatches 98; Indels 67; Gaps 11;

QY 1 NMRGVPFRLHLVLTALLPATQGNKVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 NMRRIYFQCLLVLTALLPATWTKTYRGKAGAVELPCOSSQSKRSVFMKGANQVK 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 ILGNQ---SFLYKPSKLNDRADRSRLMDQGNFPLIKNLTIEDSDTYICEVDOKE 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 ILGNQSSSSSFWLKGNSPLSNRVESKKMMQDGSFPLVIKDLRMDDSGTICEVDKGM 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 117 EYQVLVFGLTANSDFHLQGSGLTTLTSSPSSSVQCRSPRKNIGGKTLVSQLEL 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 EVELLVFRLTANPNRLLHGSGLTLTLGSPVSGVSPQMKSEPKILIEGPTCSMPKRL 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 QDSGTWTCIV-LQNKVYEFKIDIVCPAPBPKSCDKTHTCPELGSPVFLPPPKXDT 235
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 QDSGTWSCHLSFQDQNKLELDIKIIVLGFPKSA-----TYKKEGEVSEFSF----- 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 236 LMISRTPEVTCVVDVSHEDPEV--KFNMYVDGVVHNAKTKPREEOYNSTYRVSVLT 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 -----LNFEDESLSGELMWQVDGAS----- 249
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 294 LHQDVLNGKEYKCKRS-KKALPAPLEKTIISKAGQPREPQVYTLPPSRDELTKNQ 351
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 250 SAGSVSSVLEDRKVSQVQILP--DLKIQMSGLPLS--LTLPLQALHRYAGSGNLSLT 303
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 CLVKG 356
   | | |
Db 304 -LDKG 307

RESULT 32
CD4_RAT STANDARD; PRT; 457 AA.
AC P05540;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
   T4/Leu-3) (W3/25 antigen).
GN CD4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87175535; PubMed=3104900;
RX Clark S.J., Jeffries W.A., Barclay A.N., Gagnon J., Williams A.F.;
RT "Peptide and nucleotide sequences of rat CD4 (W3/25) antigen:
   evidence for derivation from a structure with four
   immunoglobulin-related domains."
RT Proc. Natl. Acad. Sci. U.S.A. 84:1649-1653 (1987).
RN (2)
RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 210-393.
RX MEDLINE=93262437; PubMed=8493535;
RX Brady R.L., Dodson E.J., Dodson G.G., Lange G., Davis S.J.,
RX Williams A.F., Barclay A.N.;

```

RT "Crystal structure of domains 3 and 4 of rat CD4: relation to the
 RT NH2-terminal domains." ;
 RL Science 260:979-983(1993).
 CC -|- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -|- SUBUNIT: Associates with p56-lck (By similarity).
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -|- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -|- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M15768; AAA0901.1; -.
 DR PIR; A27449; A27449.
 DR PDB; 1CID; 15-JUL-93.
 DR GLCOSUITEB; P05540; -.
 DR GO; GO:0042101; C-T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.
 DR GO; GO:0042289; F:MHC class II protein binding; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . . ; ISS.
 DR GO; GO:0030211; P:T-cell differentiation; ISS.
 DR GO; GO:0045058; P:T-cell selection; ISS.
 DR GO; GO:0001163; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
 DR InterPro; IPR00973; CD4_TcAg.
 DR InterPro; IPR007110; IG-like.
 DR Pfam; PF00047; Ig; 2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00403; IG; 2.
 DR PROSITE; PS00835; IG_LIKE; 1.
 DR Immnoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure.
 FT SIGNAL 1 27
 FT CHAIN 26 457 T-CELL SURFACE GLYCOPROTEIN CD4.
 FT DOMAIN 28 394 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 395 417 POTENTIAL.
 FT DOMAIN 418 457 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 28 127 IG-LIKE V-TYPE.
 FT DOMAIN 128 206 IG-LIKE C2-TYPE 1.
 FT DOMAIN 207 316 IG-LIKE C2-TYPE 2.
 FT DOMAIN 317 374 IG-LIKE C2-TYPE 3.
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 43 111 BY SIMILARITY.
 FT DISULFID 158 187 BY SIMILARITY.
 FT DISULFID 328 370 BY SIMILARITY.
 FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
 FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
 FT STRAND 213 217
 FT TURN 218 219
 FT STRAND 222 225
 FT STRAND 223 243
 FT STRAND 224 258
 FT TURN 259 260
 FT STRAND 261 265
 FT STRAND 274 275
 FT STRAND 276 278
 FT TURN 279 280
 FT STRAND 282 285
 FT HELIX 290 292
 FT STRAND 294 301
 FT STRAND 306 319
 FT STRAND 325 331
 FT STRAND 338 344
 FT TURN 345 346

FT STRAND 349 353
 FT STRAND 357 361
 FT STRAND 367 374
 FT TURN 375 376
 FT STRAND 377 385
 SQ SEQUENCE 457 AA; 51437 MW; 477BE157D30954C1 CRC64;
 Query Match 20.5%; Score 495; DB 1; Length 457;
 Best Local Similarity 32.8%; Pred. No. 3.3e-26;
 Matches 137; Conservative 57; Mismatches 132; Indels 92; Gaps 14;
 QY 1 MNRGVPFRHL-LIVLOLALLPAATGKVKYLGGKDPVELTCTASQKSIQFWKNSQ 58
 DB 1 MCRGFSFRHLPLILLQLSLKLVLTQCKTVVLGKGGASAEIPCESTRRSASFAMKSSDQ 60
 QY IKILNGGSLFTGPSKLNDRADSRSLMDQGNPLIYIKULKIEDSPTYICEVDOKREV 118
 DB 61 KTLIGYKMKLIKSLSELYSRFDSRKAMERGSPPLIINKLRBDSQTYVCELENKKEEV 120
 QY 119 QLVFGLTANSDFHLLOGSILTLTLES-PPGSSPSVQCRSPRGKNIQGKTLVSQLELQ 177
 DB 121 ELWFRVTFPGRTRLQGSLTILIDSNPKVSDPICKKSSNIVKDSKAFSTHSLRIQ 180
 QY 178 DSGTWTCTVLQNGKKEF--KIDIVPCPAPBPSPKCDKTHCPBLGGPSVPLFPKPKDT 235
 DB 181 DSGIMNCTVTLNOKKHSFDMKLSVL-----GPASTSIRAYKSEGSASFSPF----- 227
 QY 236 LMSIRTEVTCVVAVDSHEDPEVKFNMYVDGVEVHNKAKTPREBQVNSTYRVASVLTVLH 295
 DB 228 -----INLEESLQGEHRW-----KAEKAPSS----- 249
 QY 296 QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTRKNQVSLTCLVK 355
 DB 250 QSWITFSLKQNKYS-----VQKSTSNPKFQISE-----TLF-----LTLQI- 285
 QY 356 GFYPSDIAVWESNKGQPENNYKTPPVLDSDGFSLYSKLTVDKSRKQGNVSCSYW 413
 DB 286 ----PQVSLQFAGSG---NLTLT---LDR-GILYQEVNLVVMKVTOPDSTNLTLCFEM 331
 RESULT 33
 CD4_MOUSE STANDARD; PRT; 457 AA.
 ID CD4_MOUSE
 AC P06332;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
 DE T4/Leu-3) (T-cell differentiation antigen L3T4).
 GN CD4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87018845; PubMed=3094146;
 RX Tourville B., Gorman S.D., Field E.H., Hunkapiller T., Parnes J.R.;
 RT "Isolation and sequence of L3T4 complementary DNA clones: expression
 RT in T cells and brain."
 RL Science 234:610-614(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87115821; PubMed=3027575;
 RX Littman D.R., Gettner S.N.;
 RT "Unusual function in the immunoglobulin domain of the newly isolated
 RT murine CD4 (L3T4) gene."
 RL Nature 325:453-455(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Brain;
 RX MEDLINE=86152875; PubMed=3326818;
 RA Parnes J.R., Hunkapiller T.;

RT "L3T4 and the immunoglobulin gene superfamily: new relationships
between the immune system and the nervous system.";
RT Immunol. Rev. 100:109-127(1987).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=88041159; PubMed=2823269;
RA Gorman S.D., Touryvelle B., Parnes J.R.;
RT "Structure of the mouse gene encoding CD4 and an unusual transcript
in brain.";
RT Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
RL [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=98112780; PubMed=9445485;
RA Ansat-Lari M.A., Oeljen J.C., Schwartz S., Zhang Z., Muzny D.M.,
LA Lu J., Gortell J.H., Chinault A.C., Belmont J.W., Miller W.,
RA Gibbs R.A.;
RT "Comparative sequence analysis of a gene-rich cluster at human
chromosome 12p13 and its syntenic region in mouse chromosome 6.";
RL Genome Res. 8:29-40(1998).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=1247932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.L., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant P.L., Scheetz T.E.,
RA Brownstein M.J., Uedai T.B., Toohyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richard S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP SEQUENCE OF 27-43.
RX MEDLINE=86166694; PubMed=3082751;
RA Claesson B.J., Tsagaratos J., Kirszbaum L., Maddox J., McKay C.R.,
RA Brandon M., McKenzie I.F.C., Walker I.D.;
RT "The L3T4 antigen in mouse and the sheep equivalent are
immunoglobulin-like.";
RL Immunogenetics 23:129-132(1986).
RN [8]
RP DISULFIDE BONDS.
RX MEDLINE=86233454; PubMed=3086886;
RA Claesson B.J., Tsagaratos J., McKenzie I.F.C., Walker I.D.;
RT "Partial primary structure of the T4 antigens of mouse and sheep:
assignment of intrachain disulfide bonds";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4499-4503(1986).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P06332-1; Sequence=Displayed;
CC Name=2; Synonyms=Brain-specific;
CC IsoId=P06332-2; Sequence=VSP_002489;
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M36850; AAA39401.1; -
CC EMBL; M13816; AAA37267.1; -
CC EMBL; X04836; CAA28539.1; -
CC EMBL; M36851; AAA39402.1; -
CC EMBL; M17080; AAA37403.1; -
CC EMBL; M17078; AAA37403.1; JOINED.
CC EMBL; M17079; AAA37403.1; JOINED.
CC EMBL; AC002397; AAC36010.1; -
CC EMBL; BC039137; AAH39137.1; -
CC PIR; A02110; RMM574.
CC HSSP; P01730; 1WBR.
CC MGD; MGI:88335; CD4.
CC GO; GO:0042101; C-T-cell receptor complex; ISS.
CC GO; GO:0015026; Fc-receptor activity; ISS.
CC GO; GO:0042289; Fc-MHC class II protein binding; ISS.
CC GO; GO:0006955; P:immune response; ISS.
CC GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
CC GO; GO:0030217; P:T-cell differentiation; ISS.
CC GO; GO:0045058; P:T-cell selection; ISS.
CC GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
CC InterPro; IPR000973; CD4_TCSG.
CC InterPro; IPR007110; Ig-Like.
CC InterPro; IPR003596; Ig_V.
CC Pfam; PF00047; Ig_2.
CC SMART; SMART; CD4TCANTIGEN.
CC SMART; SM00406; IGV; 1.
CC PROSITE; PS00835; IG_LIKE; 1.
CC Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
CC Immune response; Repeat; Signal; Lipoprotein; Palmitate;
CC Alternative splicing.
CC FT SIGNAL 1 26
FT CHAIN 27 457 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 27 394 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 395 417 POTENTIAL.
FT DOMAIN 418 457 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 27 128 IG-LIKE V-TYPE.
FT DOMAIN 129 207 IG-LIKE C2-TYPE 1.
FT DOMAIN 208 317 IG-LIKE C2-TYPE 2.
FT CARBOHYD 318 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 42 112
FT DISULFID 159 188
FT DISULFID 328 370
FT LIPID 418 418 S-palmitoyl cysteine (by similarity).
FT LIPID 421 421 Missing (in isoform 2).
FT VARSPLIC 1 240 /FTId=VSP_002489.
SQ SEQUENCE 457 AA; 51296 MW; 1B1DA7527CB00F33 CRC64;
Qy Best Local Similarity 19.7%; Score 475; DB 1; Length 457;
Qy Matches 104; Conservative 32; Mismatches 55; Indels 6; Gaps 5;
Qy 1 NNRGVFRR-LTLVLQALLPATQGNRVVLGKKDVTVELTCTASQKSIQFMNNSQI 59
Qy 1 MCRATSLRRLILLQLLSQLAVTQKTLVLGKESASLPCSSQKQITVFTWFSQOR 60
Qy 60 KILNQG-SFLTKG--PSKLDNRADSRSLWDQGFPLIKLKIKEDSDTYICEVEDQKE 116
Qy 61 KILGHHGKQVLLRGSSPSQF--DRFDSKKGAWEKGFPLINKLXKEDSQTYICELENKE 119
Qy 117 EVQLAVFGITANSDFHLQGGSLTLTLSS--PGSSPSVQCRSPRKNIQGGKTLSSVQLE 175

```

Db      120 EVELNVEKVTSPGTSLLQGSLLTLTLDSNKNVNPLETECHKKGKGVSGSKVLSMSMLR 179
Oy      176 LQDSGWTCTVLQONQK 192
       : ||| | ||| : |||
Db      180 VQDSDFNCTVTYLDQK 196

RESULT 34
EPC_MOUSE STANDARD: PRT: 421 AA.
AC P06336: P01856: (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE I9 epsilon chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=84236092; PubMed=6329728;
RA Ishida N., Ueda S., Hayashida H., Miyata T., Honjo T.;
RT "The nucleotide sequence of the mouse immunoglobulin epsilon gene:
RT comparison with the human epsilon gene sequence.";
RL EMBO J. 1:1117-1123(1982).
[2]
RN REVISIONS.
RA Honjo T.;
RL Submitted (APR-1986) to the EMBL/Genbank/DBJ databases.
RN (3)
RP SEQUENCE OF 34-421 FROM N.A.
RX MEDLINE=83117774; PubMed=6818553;
RA Liu F.-T., Albrandt K., Sutcliffe J.G., Katz D.H.;
RT "Cloning and nucleotide sequence of mouse immunoglobulin epsilon
RT chain cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:7852-7856(1982).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X01857; CAA25977.1; -.
DR EMBL, X01857; CAA25978.1; -.
DR PIR, A02144; EHMS.
DR PIR, A02145; EHMS.
DR HSSP, P01854; 11GE.
DR InterPro; IPR007110; I9-1Ike.
DR InterPro; IPR003597; I9_C1.
DR InterPro; IPR003006; I9_MHC.
DR Pfam; PF00047; I9; 4.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KV Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN 1 90
FT DOMAIN 91 197
FT DOMAIN 198 304
FT DOMAIN 305 421
FT DISULFID 23 75
FT DISULFID 121 180
FT DISULFID 226 285
FT DISULFID 330 392
FT CARBOHYD 43 43
FT CARBOHYD 72 72
FT CARBOHYD 84 84
FT CARBOHYD 95 95
FT CARBOHYD 166 166
       N-LINKED (GLCNAC. . .) (POTENTIAL).
       N-LINKED (GLCNAC. . .) (POTENTIAL).
       N-LINKED (GLCNAC. . .) (POTENTIAL).
       N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 421 AA; 47320 MM; 8F909E1F30A06B47 CRC64;

Query Match 16.1%; Score 388; DB 1; Length 421;
Best local similarity 28.3%; Pred. No. 4,7e-19;
Matches 119; Conservative 77; Mismatches 138; Indels 86; Gaps 20;

Oy 38 ELTCTASQKXSIQPHMKNQIKLGNQGSFLTGTGPKLNDRAHSRLMDQGFPLILK 97
Db 55 ELKVTTSQVTS---WKSXAK---NFTCHVTHPFSFNSRT-----TLVR 92
Oy 98 NLIKEDSDTYI---CEVEDQKEVQV--LVFGTLAN--SDTHLLQGSLLTLTLESPGS 149
Db 93 PVTNTEPTLHLHSCDPNMFHSTIQYXCITYGHILNDVSVNMLMDREITDTL----- 146
Oy 150 SPVSQCRSPRGKNIQGGKTLISVQLELDQSGTWTCTVLQONQKVEFKIDIVPCPAPPKS 209
Db 147 AQTVLIKE--EGKLASTCSKLNITEQWMSSESTFCVK--TSQGVLYLAHTRRCPDHPR- 202
Oy 210 CDKHTHCPPELLGFSVFLPFPKPKDTLMISRTPTVCVVVDV--SHEDPEYKEN----- 261
Db 203 -----GVTVLIPPSPLD--LYQNGAPKLTCLVLDLSEKVNVTWNOEKRTSV 249
Oy 262 ----WYDGVVHNAKTPREEDQNSTYRVSVYLVTHQDMLNGEKYCKVSKALPAPI 317
Db 250 SASQWY---TKHNH-----NATTSITSLIPVAKMDISGYQCIYVDHDFPKPI 296
Oy 318 EKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNGOPENN-- 375
Db 297 VRSITKTPGQSRASVEVYFPPPEEE--SEDRKTYTCLIQNFPEPDISVQWLDGDKLISNQ 355
Oy 376 YKTPPLVLDSDG---FELYSKLTVDKSRVQGGVFCSMVHEALHN--HTQKSLISPG 431
Db 356 HSTTTP--LKSNGSNGQGFIPFSRLVAKTLWTQRKQFCQVTHALQKPRKLEKTIISTSLG 414

RESULT 35
MUC_HUMAN STANDARD: PRT: 454 AA.
ID MUC_HUMAN
AC P01871;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE I9 mu chain C region.
GN IGHM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE OF 1-434 FROM N.A.
RX MEDLINE=90332450; PubMed=2115996;
RA Friedlander R.M., Nussenzweig M.C., Leder P.;
RT "Complete nucleotide sequence of the membrane form of the human Igm
RT heavy chain.";
RL Nucleic Acids Res. 18:4278-4278(1990).
RN (2)
RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN GAL).
RX MEDLINE=75059123; PubMed=4803843;
RA Watanabe S., Barnikol H.U., Horn J., Bertram J., Hilschmann N.;
RT "The primary structure of a monoclonal Igm-immunoglobulin
RT (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-
RT type), subgroup H III. Architecture of the complete Igm-molecule.";
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509(1973).
RN (3)
RP REVISIONS (GAL).
RX MEDLINE=81066716; PubMed=6777162;
RA Mhaesco E., Barnikol-Watanabe S., Barnikol H.U., Mhaesco C.,
RA Hilschmann N.;
RT "The primary structure of the constant part of mu-chain-disease

```



```

CC expression of isoform Membrane-bound to isoform Secretd;
CC Name=Membrane-bound;
CC IsoId=P01873-1; Sequence=Displayed;
CC Name=Secretd;
CC IsoId=P01872-1; Sequence=External;
CC -1- MISCELLANEOUS: The sequence of residues 1-409 is assumed to be
CC identical with the corresponding region of the secreted form.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V00821; CAA24202.1; -
CC PIR; A02167; MHMSM.
CC HSSP; P01857; 1FC1.
CC InterPro; IPR007110; Ig_1like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.
CC Pfam; PF00047; Ig_4.
CC SMART; SM00407; IgC1; 2.
CC PROSITE; PS50835; IG_LIKE; 4.
CC PROSITE; PS00290; IG_MHC; 3.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
CC Alternative splicing; Transmembrane.
CC NON_TER 1 1
CC DOMAIN 1 105 CH1.
CC FT 106 217 CH2.
CC FT DOMAIN 218 324 CH3.
CC FT 325 436 CH4.
CC FT TRASNEM 456 473 POTENTIAL.
CC FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
CC FT DISULFID 28 89 BY SIMILARITY.
CC FT DISULFID 136 199 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
CC FT DISULFID 216 246 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
CC FT DISULFID 246 305 BY SIMILARITY.
CC FT DISULFID 293 293 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
CC FT DISULFID 353 415 BY SIMILARITY.
CC FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
CC FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
CC FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .).
CC FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .).
CC SEQUENCE 476 AA; 52656 MW; 8D476575A5204071 CRC64;

Query Match 15.5%; Score 373.5; DB 1; Length 476;
Best Local Similarity 25.2%; Pred. No. 5,2e-18;
Matches 121; Conservative 87; Mismatches 157; Indels 115; Gaps 23;

QY 37 VELTCTASQ--KKSIOFHKNKSNQIKIINGSGFLTKGPKSKLNDADRSRSLMDQGNF-- 92
DB 24 VAMGLARDPLPSTISFTNMYONTTEVIGGIRTFPT-----LRTGGKFLA 68
QY 93 ---PLIKLKIEDSDTY-ICEVEDQKEVQLVFGLTANSPDTHL---LOGOSITLTLE 144
DB 69 TSOVLSPRSILSGSDVYLKIH-----YG-GKNRLHPPIPAVAMENNVAVF 117
QY 145 SPP-----GSSP---SVQCR-----SPR-----GKNIQG----- 166
DB 118 VPRDGFSGPAPRKSGLICEATNTPKPIVSWLKDGLVSGFTDPTIENKSGTPTQ 177
QY 167 ----KTLISOLELDSGTWCTVQONQKVEFKIDIPCAPAPKSKCDKHTCELLIG 222
DB 178 YKVIITLTITSEIDWMLNVYTCRV--DHRGLTFLKNVSTCAASFST-----DIL-- 225
QY 223 PSVFLFPPKPDITMISRTPEVTCVAVVDVSHEDPEVKKNMYVDGVEVNAKTKPREEOYN 282
DB 226 --TFTIPSPAD-IFLSKANLCLVSNLAYE-ILNLSMSQSEPLETKIKIMESHFN 281
QY 283 STYRVASVLTVLHQMVLNGKEYKCVSKNKAIPAKTIKTSKAKGQPR-----POVYTL 337

```

```

DB 282 GTFSAKGVASVCEVDMMNRKEFVCTVTHRDLPSPQKKFISK----PNEVHKHPPAVYLLP 337
QY 338 PSRDEL-KNOVSLTCLVKGPPSPDIAMVEMSGO--PENNYKTPPVLD--SGSPEFLY 392
DB 338 PARQQLLRSAATVTCVKGPPSPADISVQWLQKQLLPQKRYTSAPPEPGAGFFYTH 397
QY 393 SKLVDSRRKQGVNPSGSVMHEALHNHYOKSLSPGLQDPTCAEADGELDGLMTT 452
DB 398 SILVTEENNSGTYTCVVGHEALPHLVTERTVDS-----TEGVNADEBEPENLMTT 452

RESULT 37
EPC_RAT STANDARD; PRT; 429 AA.
AC P01855;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig epsilon chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A. (IMMUNOCYTOMA IR2).
RC STRAIN=LOU/C/MSL;
RX MEDLINE=83064537; PubMed=6292865;
RA Hellman L., Petersson U., Engstroem A., Karlsson T., Bennich H.;
RT "Structure and evolution of the heavy chain from rat immunoglobulin
RT E.";
RL Nucleic Acids Res. 10:6041-6049(1982).
RN [2]
RP SEQUENCE OF 168-342 FROM N.A. (MYELOMA IR162).
RX MEDLINE=83182019; PubMed=6820340;
RA Kindsvogel W.R., Reddy E.P., Moore J.M., Faust C.H. Jr.;
RT "A cloned cDNA probe for rat immunoglobulin epsilon heavy chain:
RT construction, identification, and DNA sequence.";
RL DNA 1:335-343(1982).
RN [3]
RP SEQUENCE OF 205-306 FROM N.A.
RX MEDLINE=82174576; PubMed=6802387;
RA Hellman L., Petersson U., Bennich H.;
RT "Characterization and molecular cloning of the mRNA for the heavy
RT (epsilon) chain of rat immunoglobulin E.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1264-1268(1982).
CC -1- SIMILARITY: Contains 4 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J00744; AAA41379.1; ALT_INIT.
CC PIR; A93442; EHRT.
CC HSSP; P01854; 1TGE.
CC InterPro; IPR007110; Ig_1like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.
CC Pfam; PF00047; Ig_4.
CC SMART; SM00407; IgC1; 1.
CC PROSITE; PS50835; IG_LIKE; 4.
CC PROSITE; PS00290; IG_MHC; 3.
CC Immunoglobulin domain; Immunoglobulin C region; Repeat.
CC NON_TER 1 1
CC DOMAIN 1 89 IG-LIKE 1.
CC FT 103 201 IG-LIKE 2.
CC FT DOMAIN 205 305 IG-LIKE 3.
CC FT 314 414 IG-LIKE 4.
CC FT DOMAIN 168 168 R -> N (IN REF. 2).
CC FT CONFLICT 308 308 P -> L (IN REF. 2).

```

[illegible]

Query Match	Best Local Similarity	Matches	Score	DB	Length
15.3%	23.8%	79	369.5	1	479
Pred. No. 9.7e-18;	Mismatches 153;	Indels 161;	Gaps 23		
Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;					
Alternative splicing; Transmembrane.					
FT NON_TER	1	1	CH1.		
FT DOMAIN	1	106	CH2.		
FT DOMAIN	107	222	CH3.		
FT DOMAIN	223	327	CH4.		
FT DOMAIN	328	458	POTENTIAL.		
FT TRANSMEM	459	476	POTENTIAL (WITH LIGHT CHAIN) (PROBABLE).		
FT DISULFID	14	14	INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).		
FT DISULFID	28	90	BY SIMILARITY.		
FT DISULFID	137	200	BY SIMILARITY.		
FT DISULFID	219	219	INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).		
FT DISULFID	249	308	BY SIMILARITY.		
FT DISULFID	296	296	INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).		
FT DISULFID	356	418	BY SIMILARITY.		
FT CARBOHYD	46	46	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT CARBOHYD	114	114	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT CARBOHYD	212	212	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT CARBOHYD	261	261	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT CARBOHYD	277	277	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT CARBOHYD	284	284	N-LINKED (GLCNAC. . .) (POTENTIAL).		
SEQUENCE	479 AA;	52351 MM;	689C637A47B819FC CXC64;		
Query Match	15.3%	Score 369.5;	DB 1;	Length 479;	
Best Local Similarity	23.8%	Pred. No. 9.7e-18;	Mismatches 153;	Indels 161;	Gaps 23
Matches 123;	Conservative 79;	Pidmatch 153;	Indels 161;	Gaps 23	
21	AATGCKKVVGLKGGDTVELTCTASQ--KKSQGFHM--KNSNQIKLNGQSTLTGPGSKL	76			
17	ALTGGLNVAMG-----CLARDFLPSSVTFMSFNNSEI-----	50			
77	NDRADSRRLSDGQNFPLIKNLK-----IEDSDTY-ICEVEDQKEEVOVL	121			
51	-----SSTTV---RTFPVVRKGGDKMATSQVLVPSKDVLCQGTETELVCKVGHNSNNRDLR	102			
122	VFGLTANSDTHLQGSQSLTTLTLESPGSSPSVQCR-----	159			
103	V-----SEPPVSELPENNVSVFIPPDSPSGSGTRKSRLLICQATGFPSPKQI	147			
160	-----GKNIQGG-----KTLVSQLELQSGTW-----TCTV-----L	187			
148	SVSWLRDQKQKVESGVLTPVPAETKAGCAPATFSSMTLTRESMDLSQSLTYTCRVADHGI	207			
188	QNOKKVEFKDIVPCPAPEPKSCDHTHCPELQGPSVFLPPEPKQDTLMTSRTEVTCV	247			
208	FFDKVMSMSSECSSTPSP-----GIQVFPIAPSPADT-PLSKSARLICL	250			
248	VVDVSHPEPEKFMWYVDGVEVHNAKT-----KREBEQNSTYRVSVLYTLHQWMLNG	302			
251	VTDLITTYG-SLNISM-----ASHNGKALDTHMNITESHPNATFSAMGASVCAEDWESGE	304			
303	EYCKKVSNKALPAPIEKTISKAKQCPRE-POVYTLPPSRDELDT-KNQVSLTCLVYGFPYS	360			
305	QFTCTVTHADLPFLPKHTISKSREYAKHPRAVYVULPAPREQLVLRSAITVLCVAGFSPA	364			
361	DIAVESWESNQP--ENNYKTTPEVLD--SDGSFPLYSLTYVCKSRMQGNVFSQVMEIDA	416			
365	DVFPVQMOQRGQPLSDSKVYTSAPAPPEQAPALYFHSHTLVTBEMNNGEFTFCVVGHEA	424			
417	LHNHYTQKSLSPGLQDDETCAEAOSELDGLMTT	452			
425	LPHVYTERTVDKS-----TBEVGAEGBEGFENLWTT	455			
RESULT 39					
MUC CANFA	STANDARD;	PRT;	450 AA.		
ID MUC CANFA	201874;				
AC	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	15-JUL-1999 (Rel. 31, last annotation update)				
DE	Ig mu chain C region.				
OS	Cantis familyIaric (Dcc)				

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Flesipedia; Canidae; Canis.
RN [1]
RP SEQUENCE OF 1-177 (MOO).
RX MEDLINE=8007682; PubMed=117299;
RA "McClumber L.J., Capra J.D.;
RT "The complete amino-acid sequence of a canine mu chain.";
RN Mol. Immunol. 16:565-570 (1979).
[2]
RP SEQUENCE OF 178-450 (MOO).
RX MEDLINE=78180587; PubMed=653360;
RA Wasserman R.L., Capra J.D.;
RT "Amino acid sequence of the Fc region of a canine immunoglobulin M;
RT interspecies homology for the Igm class.";
RN Science 200:1159-1161 (1978).
DR PIR: A83131; MHDG.
DR HSSP: P01857; 1FCL.
DR InterPro: IPR007110; IG-1like.
DR InterPro: IPR003597; IG_C1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig: 4.
DR SMART: SM00407; Igcl: 2.
DR PROSITE: PS50835; IG_LIKE; 4.
DR PROSITE: PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1
SQ SEQUENCE 450 AA; 48895 MW; 9DA60DA9D1012F5D CRC64;

Query Match 15.3%; Score 369; DB 1; Length 450;
Best Local Similarity 24.1%; Pred. No. 9.7e-18;
Matches 119; Conservative 80; Mismatches 139; Indels 156; Gaps 20;
QY 36 TVELCTASQ--KKSIOFMKNSNOKILGNQSFLLTGPSKLNDRADSRSSIMDQGF 93
DB 22 TVAMGLARDPLPGSITFSWKTEBLSAINSTRG-----FP 56
QY 94 LIKKLK-----IEDSTYI-CEVE-----DOKEVOLVFGHTANSDTLHL 134
DB 57 SVLRGKXVAVTSQVFLPSVDIIQGTDEHIVCKVRHSBGBKQKBPVLPV----- 105
QY 135 QGQSILTLLESPGSSPSVQCR-----SPR-----GKNIQGSKT 168
DB 106 -----LTL--PPEVSGFIPPRDAFPGBRKSQLCOAGSFPQVMSLRDQKQIESGVT 157
QY 169 LSVQLELDQSG-----TWTC-----TVLQNKQKVEKIDIV 200
DB 158 TNEVZAKZSGPTTYKVTSMLTIOEDAMLSQSVTCVHRGLTFQONASSM----- 210
QY 201 PCPAPKSCDKHTCPPELLGSPVFLPPPKKDTLMTSRPEVTVCVVVDVSHEDPEVKF 260
DB 211 -CTSQGPV-----GISFTIPPS-FASIFMTKSAKLSCLVTLDTATYD-SVTI 254
QY 261 NMYVDGVEVNAKTPREBOYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEK 320
DB 255 SSTRENGALKHTHTNISSEHPNGITPSAMGEAVVCEEWESGEQFCCTVTHDLPSVLKOT 314
QY 321 ISKAGG-QPREQVYTLPSRDEL-TKNQVSLTCLVKGFPYSDIADVEMESNGQ--PENNY 376
DB 315 ISRPQVAVHMPVSIVLPSPREQLDRSATLSCLVLTGSPDVVQVQKQPPVPSY 374
QY 377 KTTPEVLN--SDGSFFLYSKLTVDKSRMQQGVNFSQVMEALNHNYSLSLSPG--- 431
DB 375 VTSAMPEPQAGLFAFASILTVSEEMNAGETTCVVAHESLPVRVTERVDVSTGKPT 434
QY 432 -----LQIDETCAE 440
DB 435 LYNVSLVSLDTAGZ 448

RESULT 40
MUCB_HUMAN STANDARD; PRT; 391 AA.
ID MUCB_HUMAN

AC P04220;
DT 20-MAR-1987 (Rel. 04, Created)
DR 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig mu heavy chain disease protein (BOT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN NCB1_Taxid=9606;
[1]
RP SEQUENCE.
RX MEDLINE=84184186; PubMed=6425189;
RA Barnikol-Watanabe S., Mihaesco E., Mihaesco C., Barnikol H.U.,
RA Hilschman N.;
RT "The primary structure of mu-chain-disease protein BOT. Peculiar
RT amino-acid sequence of the N-terminal 42 positions.";
RL Hoppe-Seyler's Z. Physiol. Chem. 365:105-118 (1984).
CC -1- MISCELLANEOUS: This protein has no V region homology or CH1
CC region.
DR PIR: A02163; MHUBT.
DR HSSP: P01857; 1FCL.
DR GO: GO:0005624; C:membrane fraction; NAS.
DR GO: GO:0003823; F:antigen binding; TAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-1like.
DR InterPro: IPR003597; IG_C1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig: 3.
DR SMART: SM00407; Igcl: 2.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region.
FT DOMAIN 1 42
FT DOMAIN 43 155 CH2.
FT DOMAIN 156 261 CH3.
FT DOMAIN 262 391 CH4.
SQ SEQUENCE 391 AA; 43057 MW; 9100843AF0CF021A CRC64;

Query Match 15.1%; Score 365.5; DB 1; Length 391;
Best Local Similarity 28.0%; Pred. No. 1.4e-17;
Matches 105; Conservative 65; Mismatches 132; Indels 73; Gaps 15;
QY 110 EVED--QKEEVOVLVGLTANSPTLHLLQGQSILTLTLESPPGSSPSVQCR----- 156
DB 18 EADRLIKKEEARL-----SGRD--MQVTSQPVIAELPPKVSVFVPRDGFPGNPKRS 68
QY 157 -----SPR-----GKNIQGKTLVSQLELDQSG-----TWCTYLDN--- 189
DB 69 KLICQATGFSPPROILEVSWLRGKQVSGVTTDEVEAKESGPTTYKVTSTLTIKESDWL 128
QY 190 -QKKVEFKIDIVPCPAPKSCDKHTCPPELLGSPS-----VLPFPKPKDTLMISRT 242
DB 129 GQSMFTGRVDRHGLTFPQANSS-----MCPDQDDTAIRPAIIPS-FASIFLTKST 178
QY 243 EVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTPREBOYNSTYRVVSVLTVLHQDWLNGK 302
DB 179 KLCTVLTDLTYD-SVTISWTRQDGEAVKTHTNISSEHPNATFSAVGEASICEDDMDSGE 237
QY 303 EYKKSNNKALPAPIETIKAKGP-REPQVYTLPSRDEL-TKNQVSLTCLVKGFPY 360
DB 238 RFTCTVTHDLPSPKQITIRPKGVALLRPDVLPLPARQQLMRESATITCLVTGSPFA 297
QY 361 DIAVESNGQP--ENNYKTPVLN--SDGSFFLYSKLTVDKSRMQQGVNFSQVMEALNHN 416
DB 298 DVFQMMQROQPLSPKTYTSAPMPPEQAGRFASILTVSEEMNTGETTYTCVVAHEA 357
QY 417 LNNHYTOKSLSLSPG 431
DB 358 LPRKVTERTVDKSTG 372

RESULT 41
MUC_MOUSE
ID MUC_MOUSE

```

ID MUC_MOUSE STANDARD; PRT; 455 AA.
AC P01872;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig mu chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RX MEDLINE=81076590; PubMed=6255422;
RA Kawakami T., Takahashi N., Honjo T.;
RT "Complete nucleotide sequence of mouse immunoglobulin mu gene and
RT comparison with other immunoglobulin heavy chain genes.";
RN Nucleic Acids Res. 8:3933-3945(1980).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=82051295; PubMed=6795090;
RA Goldberg G.I., Vanin E.F., Zrolka A.M., Blattner F.R.;
RT "Sequence of the gene for the constant region of the mu chain of
RT Balb/c mouse immunoglobulin.";
RN Gene 15:33-42(1981).
RL [3]
RP SEQUENCE FROM N.A. (MVELOMA TEPCL83).
RX MEDLINE=81165562; PubMed=6260591;
RA Auftray C., Rougeon F.;
RT "Nucleotide sequence of a cloned cDNA corresponding to secreted mu
RT chain of mouse immunoglobulin.";
RN Gene 12:77-86(1980).
RL [4]
RP SEQUENCE (MVELOMA PROTEIN MOPC 104E).
RX MEDLINE=79223904; PubMed=111247;
RA Kehry M.R., Sibley C.H., Fuhrman J.S., Schilling J.W., Hood L.E.;
RT "Amino acid sequence of a mouse immunoglobulin mu chain.";
RN Proc. Natl. Acad. Sci. U.S.A. 76:2932-2936(1979).
RL [5]
RP REVISION (MOPC 104E).
RX MEDLINE=83075344; PubMed=6816276;
RA Kehry M.R., Fuhrman J.S., Schilling J.W., Rogers J., Sibley C.H.,
RA Hood L.E.;
RT "Complete amino acid sequence of a mouse mu chain: homology among
RT heavy chain constant region domains.";
RN Biochemistry 21:5415-5424(1982).
RL -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=During differentiation, B lymphocytes switch from
CC expression of isoform Membrane-bound to isoform Secreted;
CC Name=Secreted;
CC IsoId=P01872-1; Sequence=Displayed;
CC Name=Membrane-bound;
CC IsoId=P01873-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, V00827; -, NOT_ANNOTATED_CDS.
DR PIR, A02166; MIMS.
DR HSSP, P01857; IFC1.
DR InterPro, IPR007110; IG-1like.
DR InterPro, IPR003597; IG_c1.
DR InterPro, IPR003006; IG_MHC.
DR Pfam, PF00047; Ig_4.
DR SMART, SM00407; Igcl, 2.
DR PROSITE, PS00835; IG_LIKE; 4.
DR PROSITE, PS00290; IG_MHC; 3.

```

```

KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing.
FT NON_TER 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 217 CH2.
FT DOMAIN 218 324 CH3.
FT DOMAIN 325 436 CH4.
FT DOMAIN 437 455 C-TERMINAL SEGMENT OF SECRETED FORM.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 28 89 BY SIMILARITY.
FT DISULFID 136 199 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 216 216 BY SIMILARITY.
FT DISULFID 246 305 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 293 293 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 323 415 BY SIMILARITY.
FT DISULFID 454 454 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .).
FT VARIANT 78 78 S -> N (IN MOPC 104E).
FT VARIANT 101 101 H -> Q (IN MOPC 104E).
FT VARIANT 226 226 T -> N (IN TEPCL83 AND MOPC 104E).
FT VARIANT 258 258 N -> T (IN MOPC 104E).
FT VARIANT 258 258 N -> S (IN TEPCL83).
FT VARIANT 358 358 L -> K (IN TEPCL83 AND MOPC 104E).
SQ SEQUENCE 455 AA; 50101 MW; 4CB57CB602F9B51 CRC64;

Query Match 15.0%; Score 362; DB 1; Length 455;
Best Local Similarity 25.3%; Pred. No. 2,9e-17;
Matches 116; Conservative 84; Mismatches 149; Indels 110; Gaps 22;

QY 37 VELTCTASQ--KKSIGFTWKNSNQIKILGNQSGFLTKPSPKLNDRADRRSLMDQNF-- 92
DB 24 VAMGLIADFLPSPSTISFTWNYQNTEVIGQIRTEPT-----LRTGAKYLA 68
QY 93 ---PLIINKIKIEDSDY-ICEVEDQKEEVQVLVRLTANSPTHL-----LQOQSILTLE 144
DB 69 TSQVLLSPKSIDSGDEYLVCXIR-----YG-GRNDLHVPIPAVAMENNVAVF 117
QY 145 SPP-----GSSP----SVQCR-----SPR-----GKNIOGQ----- 166
DB 118 VPRRGFGSPAPRKSKLICEATNFTPKPIYTWMLKDGGLVSGFTTDDVTIENKSTQOT 177
QY 167 ---KTLVSQLELDQSGTWTCTVILQNOQKVEFKIDIVCPAPPEKSCDKHTTCPELLGG 222
DB 178 YKVISTLTITSEIDMLNLNVYTCRV--DHRGLTFLKNVASTCAASPST-----DIL-- 225
QY 223 PSVFLFPPKPKPDYLMISRTPTVYVDSHEDPEVKKNVYVDGVEVNAKTKPREQYN 282
DB 226 --TFPIPSPAD-IFLSKSNMLTCLVSNLAYE-TLNISSMSQSGPLETKIKIMESHFN 281
QY 283 STYRVVSLVTLVHODMLNGKEYCKVSNKALPAPIEKTISKAKGQPRB-----POVYTLF 337
DB 282 GFISAKGVASVCEVEMNNRKEFVCTVTHRDLPSPQKRTISK-----PNEVHNGPRAVYLLP 337
QY 338 PSRDELTL-KNOVSLTCLVKGFPYSDIAVWESNQ--DENNYKTTTPVLD--SDGSFFLY 392
DB 338 PAREQLNRESATVTCVLKVGFSPADISVQMLQROGLLPOEKVYTSAPMPERGAPGFYFTH 397
QY 393 SKLTVDKSRMOQGNVFGSVVHEALAHNYTQKSLSLSG 431
DB 398 SILTVTEEMNSGETYTCVVGHEALPHLVTERVYDKSTG 436

RESULT 42
EPC_HUMAN
ID EPC_HUMAN STANDARD; PRT; 428 AA.
AC P01854;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```



```

FT STRAND 377 378
FT TURN 380 381
FT STRAND 384 385
FT STRAND 387 389
FT HELIX 394 398
FT TURN 399 400
FT STRAND 404 408
FT STRAND 418 421
SQ SEQUENCE 428 AA; 47019 MW; 25C4CA072AA558A0 CRC64;

Query Match 15.0%; Score 361.5; DB 1; Length 428;
Best Local Similarity 28.9%; Pred. No. 2.9e-17;
Matches 102; Conservative 53; Mismatches 137; Indels 61; Gaps 10;

QY 90 GNPPLIKLKIEDSTYICEVEDQEEVQLVFGITAN--SDTHLQGSLLTLTLESP 147
DB 125 GHPPTIQL-----CLVSGTPTGTITITWLEQVNDVDL----- 160
QY 148 GSSPSVQCSPPGKNIQGGKTLVSQLELDQSGTWTCTVLQNKVFKIDIVPCPAP 207
DB 161 ----STAATQEGELASTOSELTLSQKMLSDRTTYCQVTTQGHTE----- 203
QY 208 KSCDKTHTCPDL-LGGPSVFLFPPPKPDITMISRTPEVTCVVVDVSHEDPEVKFMYVDG 266
DB 204 ----DSTFKCADSNPRGVSAVLSRSPFD-LFIRKSPITICLVVDLAPSKGTNLTWSRAS 259
QY 267 VEHNAKTPREQVNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKITSKANG 326
DB 260 GKVHNSHTEKEKQNGTLVTSTLPVGTQDMIEGTTCQVTHPHLPALMRSTTKTSG 319
QY 327 QPREPOVYTL-----PPSRDELTKQVSLTCLVKGFPYSDIVVEMESN--GGPENNYKTT 379
DB 320 PRAAPEVVAFAITPEWGSRDKRT-----LACLIONPMDISVQMLHNEVQLPDARHSTT 374
QY 380 PVLSDSGSFPLYSKLTVDKSRMQGVNFGSCVMEHALNHYT-QKSLSLSPG 431
DB 375 QPRKTGSGFFVFSRLVETRAEWQKDEFTICRAVHEAASPQTQGAIVSVNG 427

RESULT 43
MUC_RABIT STANDARD; PRT; 458 AA.
ID MUC_RABIT STANDARD; PRT; 458 AA.
AC P03988;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DI Ig mu chain C region secreted form.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NC NCBI_TaxID=9986;
RN (1)
RP SEQUENCE FROM N.A. (A2 ALLOTYPE).
RX MEDLINE=84088930; PubMed=6418803;
RA Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.;
RT "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain
RT of Vha2 allotype: comparisons with Vha1 and membrane mu sequences."
RL J. Immunol. 132:490-495 (1984).
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing: Named isoforms=2;
CC Comment-during differentiation, B lymphocytes switch from
CC expression of isoform Membrane-bound to isoform Secreted;
CC Name=Secreted;
CC Name=Membrane-bound;
CC Name=IsoidaP03988-1; Sequence=Displayed;
CC Name=IsoidaP04221-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@ib-sib.ch).
CC -----
CC EMBL; K01357; -, NOT_ANNOTATED_CDS.
CC PIR; A02164; MHRB.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR007110; IG-1like.
CC InterPro; IPR003597; IG_C1.
CC InterPro; IPR003006; IG_MHC.
CC Pfam; PF00047; Ig; 4.
CC SMART; SM00407; IGc1; 2.
CC PROSITE; PS50835; IG LIKE; 4.
CC PROSITE; PS00290; IG_MHC; 3.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
CC Alternative splicing.
CC FT NON_TER 1
CC FT DOMAIN 1 106 CH1.
CC FT DOMAIN 107 222 CH2.
CC FT DOMAIN 223 327 CH3.
CC FT DOMAIN 328 458 CH4.
CC FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
CC FT DISULFID 28 90 BY SIMILARITY.
CC FT DISULFID 137 200 BY SIMILARITY.
CC FT DISULFID 219 219 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
CC FT DISULFID 249 308 BY SIMILARITY.
CC FT DISULFID 296 296 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
CC FT DISULFID 356 418 BY SIMILARITY.
CC FT CARBOHYD 46 46 N-LINKED (GLUCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 114 114 N-LINKED (GLUCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 212 212 N-LINKED (GLUCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 261 261 N-LINKED (GLUCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 277 277 N-LINKED (GLUCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 284 284 N-LINKED (GLUCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 445 445 N-LINKED (GLUCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 458 AA; 49897 MW; F338D6A3B28E4864 CRC64;

Query Match 14.9%; Score 360; DB 1; Length 458;
Best Local Similarity 23.8%; Pred. No. 4e-17;
Matches 118; Conservative 76; Mismatches 145; Indels 156; Gaps 22;

QY 21 AATQGNKYVLGKKGDVETLCTASQ--KSIQFHW--KNSNQIKLQNGSFLTKGPKL 76
DB 17 ALTDGNLWAMG-----CLARDPLSPSVTFGSPFNQSEI----- 50
QY 77 NDRADSRSLMDQGNFPLITKLNK-----IEDSTY-ICEVEDQEEVQL 121
DB 51 ----SSRTV--RTFPVKKRGDKYMATSOVLVPSKVDLQGTETLVCKVQHSNMRDLR 102
QY 122 VFGULANSDFHLQGSLLTLTLESPGSSPSVQCR-----SPR-- 159
DB 103 V-----SFPVDSLELPNVSVFIPPRDSFGSGCTRKSLRLICATGFSFKQI 147
QY 160 -----GKNIQGG-----KTLVSQLELDQSGTW-----TCTV-----L 187
DB 148 SVSWMRDQGXVSGVLTKEVAETKGAIPAFTSISSMLTIESDWLSGLTCTVRDHGI 207
QY 188 QNOKVVEFKIDIVPCADPEPKSCDKHTCPPELLGGPSVFLFPPPKDITMISRTPEVTCV 247
DB 208 FFDKVVSSSSSECSSTPSP-----GIQVFPPIAPSFADT-FLSKSARLITCL 250
QY 248 VVDVSHEDPEVKFMYVDVGVENHAKT-----KREEDVNSTYRVSVLTVLHODMLNGK 302
DB 251 VTDLTTYG-SLNTISV-----ASHNGKALDTHMNTTSHPNATFSAMGASVCAEDWESGE 304
QY 303 EYKCVSNKALPAPIEKITSKAKQPRE-POVYTLTPSRDELTT-KNQVSLTCLVKGFPYS 360
DB 305 QPTCTVTADLPFLPKHTISKREVAKHPRAVYVLPAPREGLVLRKSTVTCVKGSPFA 364
QY 361 DIAVWESNQP--ENNYKTTTPVLD--SDGSFPLYSKLTVDKSRMQGVNFGSCSVNHEA 416
DB 365 DVFVQMOQRGQPLSDSKVTSAPAPQALVFTHTLVTETEDMNSGTEFTCVVGHFA 424
QY 417 LHNHYTQKSLSLSPG 431

```

Db 425 LPHWTERTVKSTG 439

RESULT 44

MUC_SUNMU STANDARD; PRT; 457 AA.

AC P20768;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ig mu chain C region.
 OS Suncus murinus (House shrew) (Mus shrew).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Insectivora; Soricidae; Crocidurinae; Suncus.
 NCBI_Taxid=9378;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=89232144; PubMed=2497033;
 RA Ishiguro H., Ichihara Y., Nakatsu T., Kurosawa Y.;
 RT "Nucleotide sequence of Suncus murinus immunoglobulin mu gene and
 comparison with mouse and human mu genes.";
 RL FEBS Lett. 247:317-322(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X13920; CA32113.1; ALT_INIT.
 DR PIR, S03961; S03961.
 DR HSSP, P01842; 7EAB.
 DR InterPro; IPR007110; IG-1like.
 DR InterPro; IPR003597; IG_C1.
 DR InterPro; IPR003006; IG_MHC.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00407; IGc1; 3.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
 FT NON TER 1 1
 FT DOMAIN 1 105 CH1.
 FT DOMAIN 106 220 CH2.
 FT DOMAIN 221 326 CH3.
 FT DOMAIN 327 457 CH4.
 FT DISULFID 13 89 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
 FT DISULFID 27 133 BY SIMILARITY.
 FT DISULFID 136 200 BY SIMILARITY.
 FT DISULFID 217 217 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
 FT DISULFID 248 307 BY SIMILARITY.
 FT DISULFID 295 295 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
 FT DISULFID 355 417 BY SIMILARITY.
 FT DISULFID 456 456 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 457 AA; 50074 MW; 56C8C086DA4462E9 CRC64;

Query Match 14.7%; Score 354.5; DB 1; Length 457;
 Best Local Similarity 25.5%; Pred. No. 9.4e-17;
 Matches 122; Conservative 71; Mismatches 163; Indels 123; Gaps 19;

QY 19 LPAATGKAVVIGKGDVVELCTASQ--KKSIOGHMKNQIKLNGSGFLTGPSKL 76
 DB 17 LPDETQ-----VTGLCLARDPLPRVTFSMKFKKSSSI----- 49

QY 77 NDRADSRRLMDQGNFPLIKNLK-----IEDSPYI-CEVEDQKEVQL 121
 DB 50 -----SSQNTY---NPEVFPTGKMATSQVLLPSTALISTDYICHTKHITGKEKK 101
 QY 122 VFGILTANSDTHLLOQOSLTLTLESPPGSSPVQCRSPKQKIQGCKTLSVQLELDQSG- 180
 DB 102 V-----ELQVTELPFPVNSIFV--PFR--NSPSGNHPTSLQICQASGF 141
 QY 181 ---TWCTVLQONKQVFEKFDIVPCAPRPEKSCDKT-----HTCP 217
 DB 142 SPRTIYMSWLORGEPVPPSLVSTSAVEAEKSGGPTFRVYSRLTTENEMLSOREFTCQ 201
 QY 218 EL-----LG-----GPSVFLPPPKDLMISRTPEVTCVVVDVSHEDPEV 258
 DB 202 ALHKGLTFQKNVSSVCMGDDTSTGISVFLPPIFAN-IFLTQSAQLCLVATGLATYD-SL 259
 QY 259 KFNWYDVGVHNAKTKPREEOYNSTYRVSVLTLTQMDLNGKEYCKVSNKALPAPIE 318
 DB 260 DISWRQNGEALQTHVNISSHPNSTPTAKGHASVCREEWSGKFTCTVQHSDDLPSPLK 319
 QY 319 KTISKAKQGRRE-POVYTLPPSDEL-TKNQVSLTCLVKGFPYSDIAVEMESNGQP--EN 374
 DB 320 QSLSRPQDVANDPSPVFLPPAQQLKRESASTTCLVKQFSPDPVFTVQHHQGPVDPK 379
 QY 375 NYKTPPEVLD--SDGSFFLYSKLTVDKSRWQGNVPSGVMEALNHNHYQKSLSPG 431
 DB 380 HYTNSNTPPEQNGLVFLVHSILTVSEKMSGSSGFSFCVVGHEALPLSTVEKAVDKNSG 438

RESULT 45

MUC_MESAU STANDARD; PRT; 454 AA.

AC P06337;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig mu chain C region.
 OS Mesocricetus auratus (Golden hamster).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Scturognathu; Muridae; Cricetinae;
 CC Mesocricetus.
 NCBI_Taxid=10036;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85297761; PubMed=2994005;
 RA McGuire K.L., Duncan W.R., Tucker P.W.;
 RT "Phylogenetic conservation of immunoglobulin heavy chains: direct
 RL Nucleic Acids Res. 13:5611-5628(1985).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X02804; CA26574.1; --
 DR HSSP, P01854; 1IGE.
 DR InterPro; IPR007110; IG-1like.
 DR InterPro; IPR003597; IG_C1.
 DR InterPro; IPR003006; IG_MHC.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00407; IGc1; 2.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
 FT NON TER 1 1
 FT DOMAIN 1 105 CH1.
 FT DOMAIN 106 218 CH2.
 FT DOMAIN 219 324 CH3.
 FT DOMAIN 325 454 CH4.

```
FT DISULFID 13 13 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 27 98 BY SIMILARITY.
FT DISULFID 135 198 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 215 215 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 245 304 BY SIMILARITY.
FT DISULFID 292 292 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 352 414 BY SIMILARITY.
FT DISULFID 453 453 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 454 AA; 50196 MW; 4DA2134612BE1469 CRC64;

Query Match 14.3%; Score 344; DB 1; Length 454;
Best Local Similarity 24.9%; Pred. No. 4,7e-16;
Matches 115; Conservative 74; Mismatches 158; Indels 114; Gaps 19;

QY 37 VELTCTASQ--KKSIOFHWKNSNQIKLQNG--SFLTKGSKLNDRADSRSLMDQGNF 92
DB 23 VAMGCLARDFLPSSISFSWMYQWKESEV--NQGVTEPTL-----RMGEKXAATSQVFL 73
QY 93 PLIKNLKIEDSPTT-ICEVEDQKEVQLVFGELTANSSTHLLQGSLLTLTLESPPGSSP 151
DB 74 P--PKSVLEGSDVYLVCVHNGNTKDKLRV-----PIPVTEMNPNVSV 115
QY 152 SVQCR-----SPR-----GKNIOG----- 166
DB 116 FPPSRDAFSGPARKSRKLTCEASNFSPKQITVSMRDGKPVKSGFTTEPVTEDEKSGGR 175
QY 167 -----KTLVSQLELDQSGTWCTVQLQNKVEFKIDIPCAPRPFKSCDKHTTCPELIG 221
DB 176 TYKVIITLTITSDMLNSVTTCRV--DHRGLTFMKNVSGTCAASPT----- 221
QY 222 GGSVLPFPKPKDT-LMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTPREKQ 280
DB 222 --DIAFPPIPSFVGIFLNKSAITLTCLVNLATYD-TLNIWSRSRSGEPLFETKXLTESH 278
QY 281 YNSTRVVSVLTVLVHODMNGEYKCKVSNKALPAPIETKISKAKQPRE-----POVYT 335
DB 279 PGTGTSALCEANVCVEDMDSGKEFCVCTVTRDPSQKFKISK---PEMNKTPPAAVYQ 334
QY 336 LPPSRDEL-TKNQVSLTCLVKGFPSPDIAMWESNQGP--ENNYKTPPVLDSGS--FF 390
DB 335 QPLABQLLRRESATVTCLVKGFSPRADIYVQMLQKQPLSQKVTYSAPMRERQAPHLXF 394
QY 391 LYSKLTVDKSRWQGNVSGSVMEALHNHYTKSLISLSPG 431
DB 395 THSVLTVTEEMNSGETYTCVVGHEALPMTVTERTVDRSTG 435

RESULT 46
HVC2_HETFR STANDARD; PRT; 438 AA.
ID_P23085;
AC 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain C region (Clone 12022) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OC NCBI_Taxid=7792;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Litman R., Shamloot M.J., Litman G.W.;
"Complete structure and organization of immunoglobulin heavy chain
```

```
RT constant region genes in a phylogenetically primitive vertebrate.";
RL EMBL J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 4 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07784; CAA30617.1; -.
DR PIR; S00980; HVRK2.
DR HSSP; P01842; 7PAB.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig1; 4.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_1like; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT DOMAIN 1 5 101 IG-LIKE 1.
FT DOMAIN 108 202 IG-LIKE 2.
FT DOMAIN 213 305 IG-LIKE 3.
FT DOMAIN 315 415 IG-LIKE 4.
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 438 AA; 47904 MW; 6E67A9C5BBF7AA7B CRC64;

Query Match 12.9%; Score 312; DB 1; Length 438;
Best Local Similarity 25.4%; Pred. No. 6,4e-14;
Matches 113; Conservative 67; Mismatches 155; Indels 110; Gaps 17;

QY 23 TQGNV---VIGKSDYELTCTASQKSIQFHWKNSNQIKLQNGSFLTKGSKLNDR 79
DB 49 TGLTIYVSANLKKKT-----YRSSQLTTESE-----VGSKIYCE 86
QY 80 ADSRSLMDQGNFPLIKNLKIEDSPTTICEVEDQKEVQLVFGELTANSSTHLL----- 133
DB 87 VRRGSLM-----IKELDKDGIIVPTVYLLTQSSSEETISRRFA 126
QY 134 -----LQGSLLT-----LESPPGSSPSVQCRSPGKNIQGGKTLVSQLELD 178
DB 127 TVLCSIIDFHESITVSMKDGQPMDSGFTVSPICEVNG-----NFSATSRLLTPAGWFS 182
QY 179 SGTWCTVQLQNKVEFKI--DIVCPARPFKSCDKHTTCPELIGSPVLPFPKPKDTL 236
DB 183 NTVYTCQVAHQEVYQSRNITGSQVPCS-----IGDPVYIKLLPSTIEQVL 226
QY 237 MISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTPREKQ-----NSTRVVSVL 291
DB 227 -LEAVTVITCV---VSNAPYGVNWM-----TQEKPLKSELAVNAPGSDSVISTV 274
QY 292 TVLHODMNGEYKCKVSNKALPAPIETKISKAKQPREQV-YTLPSPSDELTKQNS 349
DB 275 NISTQAWLSGAFFYCVVSHQDLFPPLRASIHKEEVKIDREFVSVLPPPADVSAQRPLS 334
QY 350 LTCLVKGFPSPDIAMWESNQGPEN--NYKTPPVLDSG--SFLYKSLTVDKSRWQGN 406
DB 335 LTCLVGRGSPREIFIKWTVNDKSVNPGNYKTEVWAENDNRSFYLSLSIAAEWASGA 394
QY 407 VFSQSVMEALHNHYTKSLISLSPG 431
DB 395 SYSCVVGHEALPPIKINTVKKSSG 419
```

```

RESULT 47
MUC_CHICK STANDARD; PRT; 446 AA.
AC P01875;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig mu chain C region.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RN SEQUENCE OF 80-446 FROM N.A.
RX MEDLINE=83299221; PubMed=6310496;
RA Dahan A., Reynaud C.-A., Weill J.-C.;
RT "Nucleotide sequence of the constant region of a chicken mu heavy
RL chain immunoglobulin mRNA.";
RN Nucleic Acids Res. 11:5381-5389 (1983).
RN [2]
RP SEQUENCE OF 1-79 FROM N.A.
RA Weill J.-C.;
RL Submitted (MAR-1986) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X01613; CAA25762.1; -.
DR F1R; A02170; MCH.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig_1-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1
FT DOMAIN 1 105 CH1.
FT 209 CH2.
FT DOMAIN 210 316 CH3.
FT DOMAIN 317 427 CH4.
FT 428 446 C-TERMINAL REGION.
FT DOMAIN 428 446 INTERCHAIN (WITH A LIGHT CHAIN)
FT DISULFID 16 16 INTERCHAIN (WITH A HEAVY CHAIN)
FT DISULFID 27 85 (BY SIMILARITY).
FT DISULFID 135 190 BY SIMILARITY.
FT DISULFID 237 296 BY SIMILARITY.
FT DISULFID 284 284 INTERCHAIN (WITH A HEAVY CHAIN)
FT DISULFID 344 406 (BY SIMILARITY).
FT DISULFID 445 445 INTERCHAIN (WITH A HEAVY CHAIN)
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 446 AA; 48173 MW; 3CB0CE108949BD17 CR664;
Query Match 12.9%; Score 310.5; DB 1; Length 446;
Best Local Similarity 26.2%; Pred. No. 8.3e-14;

```

```

Matches 119; Conservative 67; Mismatches 179; Indels 89; Gaps 19;
QY 28 VLIGKKGDIV-ELTCTA--SQKSIQPHMKNSQIKILQNGSEFLTGSPKLANDRADSR 84
D 13 LVLCSPSDSVYVTCANAFQDPSSIAFTWDSNNSSVSGM--DVIPVIGSPYRAVSR- 69
QY 85 SLWDQGNPPIIKNLKIEDSDTYICEVEDQKEEVQLIVFG---LTNSDTHLLQGSLSLT 140
D 70 -----IQMQSESEKQPFRCRAAHPRGVNEVVMNPPIPTBN-----GIPLF 113
QY 141 LTLESP-----PGSSPSVQC-----RSPR-----GKNIGGK 167
D 114 VTMPPREDEPGFRRASILICQTRGRRPTEVTWYNGSPVAAATTAATTGVEVAES 173
QY 168 TLVSQLELDGSGTWTCTVQONQKVEFKIDIVPCAPAPKSCDKHTC---PELLGAPS 224
D 174 RISVTESEMDPTGATFSCV-----EGEARNNSKRNCEGIEPVOODIA 216
QY 225 VFLFPPPKDITLMSRPEVTCVVDVSHEDP-EVKENMYVDGVEVNAKTKPREQYNS 283
D 217 IRVITPSPFD-IFISKATLTCRVSNMVMNADGLEVSW-WKEKGKLETAIGK-RVLQSN 273
QY 284 TYRVSVTLVLIHQWMLNGKXKVKVSNKALPAPIEKTISKAK-CQPREPOVYTLPPSRDE 342
D 274 LYTVDGVAIVCASWGDGDGVCKVKNPDLFPWEERKRTKASNAAPPSSVYPPPTQ 333
QY 343 LTKNQ-VSLTCLVGFYPSDIAVWESNGO--PENNYKTPPVLD--DSFFLYSKLTV 397
D 334 LINGQRSLVTCMAGFPPHLPFRMKNGEPLPQSGVTSAPAMENPENSVAIVSLGV 393
QY 398 DKSRWQGNVFGSCVMEALHNHYTKSLSPG 431
D 394 GAEEWGAQNVYTCIVGHEALPLQACQSVDRASG 427
RESULT 48
HVCW_HETFR
ID HVCW_HETFR STANDARD; PRT; 461 AA.
AC P21088;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region, membrane-bound form (Clone 3050).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Spleen;
RC MEDLINE=88328985; PubMed=3138109;
RX Kokubu F., Hinde K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
RL constant region genes in a phylogenetically primitive vertebrate";
EMBO J. 7:1979-1988 (1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07781; CAA30614.1; ALT_SEQ.
DR PIR; S01854; HVRKO.
DR HSSP; P01842; 7PAB.
DR InterPro; IPR007110; Ig_1-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00407; IgC1; 2.

```

```

DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KM Alternative splicing; Transmembrane.
FT NON_TER 1 1
FT DOMAIN 1 99 CH1.
FT DOMAIN 100 205 CH2.
FT DOMAIN 206 308 CH3.
FT TRANSMEM 309 418 CH4.
FT CARBOHYD 438 458 POTENTIAL.
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 461 AA; 50762 MW; 029591C885DF911 CRC64;

Query Match 12.2%; Score 295.5; DB 1; Length 461;
Best Local Similarity 26.6%; Pred. No. 8.9e-13;
Matches 119; Conservative 66; Mismatches 157; Indels 105; Gaps 23;

OY 53 MKSNQIKILGNGSFLTKGPKSLNDRADSRSLMQGNPLIITKLIKIEDSDTYICEV- 111
DB 41 WKDNNEPITTG-----LKITPSVLNKKKITYS-----SOLTTSEVGSSKIT-CEVR 88
OY 112 -----EDQKEVOLLVFG--LTNSDTHLLQGSLTLT-- 142
DB 89 RGSVWIKIIPDCKGDKVHPVTILTQSSSEIRSRFPATVLCGIIDFH--PESIVSWL 145
OY 143 -----LESPPGSPSVQCSPPRGN--IQGGTLISQLELDQSGTWTCTVLONQKVEP 195
DB 146 KQGQHMSGRTVSP--C-----GVNGTFSATSRITVPAEMFNKRYTCOV--SHQGVQ 197
OY 196 KIDI-----VPCPAPREKSCDKHTCPCLLGGPSVLPFPKPXDTLMISRPEVTCVVVD 251
DB 198 SRRTSSQVPC-----SCN-----DPYIKLLPSIEVL--LEATVITLCV--V 237
OY 252 SHEDPEVKNVYDGVENVNAKTPREBOYNSTYRVSVLTVLHQMNGKEKCYKSNK 311
DB 238 SNAPYGVNVSW---TQEQSKLKSEIAVQPEDADSYISTVNISTQMLSGAEFYCVVNHQ 294
OY 312 ALPAPLEKTIKAKGP-REPOV-YTLPPSRDELTKNOVSLTLVKGFPSPDIANWESN 369
DB 295 DLETPLRASIRKEVXDLREPSYIILSPAEDVSAGRFSLTCLVNGFPPEIFVEMTVN 354
OY 370 GOPEN--NYKTPPVLDSD--GSFELYSKLTVDSRMQGNVSCVMHEALNHYTKSL 426
DB 355 DKSVPNGTKNTVMENDNSSTFYTLSLISAAEWSAGSYCVGHEAT----- 405
OY 427 SLSPGLDDETCAEADGELDLMTTD 453
DB 406 ---PLKINRTVTKSSDS- DHIMWD 428

RESULT 49
ALC2_HUMAN STANDARD; PRT; 340 AA.
ID ALC2_HUMAN
AC P01877;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 1g alpha-2 chain C region.
GN IGHA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
ON (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=84130179; PubMed=6421489;

```

```

RA Planagan J.G., Lefranc M.-P., Rabbits T.H.;
RT "Mechanisms of divergence and convergence of the human immunoglobulin
RL alpha 1 and alpha 2 constant region gene sequences.";
RL Cell 36:681-688 (1984).
RN (2)
RP SEQUENCE (BUT).
RX MEDLINE=78137069; PubMed=416441;
RA Torano A., Putnam F.W.;
RT "Complete amino acid sequence of the alpha 2 heavy chain of a human
RT IgA2 immunoglobulin of the A2m (2) allotype.";
RL Proc. Natl. Acad. Sci. U.S.A. 75:966-969 (1978).
RN (3)
RP SEQUENCE (MYELOMA PROTEIN LAN).
RX MEDLINE=79180140; PubMed=286295;
RA Tsuzukida Y., Wang C.-C., Putnam F.W.;
RT "Structure of the A2m(1) allotype of human IgA -- a recombinant
RT molecule.";
RL Proc. Natl. Acad. Sci. U.S.A. 76:1104-1108 (1979).
RN (4)
RP REVIEW.
RX MEDLINE=91054387; PubMed=2241915;
RA Kerr M.A.;
RT "The structure and function of human IgA.";
RL Biochem. J. 271:285-296 (1990).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- SUBUNIT: Monomeric or polymeric.
CC -1- MISCELLANEOUS: The sequence of the A2m(1) allotype is shown.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
CC EMBL, J00221; AAB59396.1; ALT_INIT.
CC PIR, B22360; B22360.
CC HSSP, P01810; 2PBT.
CC Genew; HGNC:5479; IGHA2.
CC MIM; 147000; -.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SMO0407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN 1 98
FT DOMAIN 112 207
FT DOMAIN 215 317
FT DISULFID 26 85
FT DISULFID 101 101
FT DISULFID 109 109
FT DISULFID 110 167
FT DISULFID 134 191
FT DISULFID 169 169
FT DISULFID 179 179
FT DISULFID 237 300
FT DISULFID 339 339
FT CARBOHYD 47 47
FT CARBOHYD 92 92
FT CARBOHYD 131 131

```

```

FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .).
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .).
FT VARIANT 93 93 P -> S (IN A2M(2) ALLOTYPE).
FT VARIANT 102 102 /FTID=VAR 003879. ALLOTYPE).
FT VARIANT 102 102 P -> R (IN A2M(2) ALLOTYPE).
FT VARIANT 279 279 /FTID=VAR 003880. ALLOTYPE).
FT VARIANT 279 279 F -> Y (IN A2M(2) ALLOTYPE).
FT VARIANT 296 296 /FTID=VAR 003881. ALLOTYPE).
FT VARIANT 296 296 D -> E (IN A2M(2) ALLOTYPE).
FT VARIANT 326 326 /FTID=VAR 003882. ALLOTYPE).
FT VARIANT 326 326 V -> I (IN A2M(2) ALLOTYPE).
FT VARIANT 335 335 /FTID=VAR 003883. ALLOTYPE).
FT VARIANT 335 335 V -> A (IN A2M(2) ALLOTYPE).
FT SEQUENCE 340 AA; 36508 MM; 989227007563276 CRC64;

Query Match 12.2%; Score 294.5; DB 1; Length 340;
Best Local Similarity 29.2%; Pred. No. 7.1e-13;
Matches 103; Conservative 50; Mismatches 143; Indels 57; Gaps 18;

OY 122 VFGLTANSQDTH-----LLOG-----QSLTTLTSPSSPSVOCRS-PRGKNIQGGKT 168
DB 8 VFPLSLDSTPDGNNVYVACLVGFFPQEPPLSVTMSB---SGQNVYARPPSSQDASGDL 64
OY 169 LSVSQLEL-----QDSGTWTCTVLONOKKVEFKDI-VCPPAPBPSCDKHTYCPBELLGG 222
DB 65 TTSSQLTLTPATQCPDGKSVTGHVKHYTNPSQ---DVTYPCVPYPPPPC-----C-----H 111
OY 223 PSVLPFPKPKDTLMISRTPEVTCVVDVSHEDPEVKRWYDGEVHNATKPREEOYN 282
DB 112 PRLSLHRALEBD-LILGSEBANLCTLTGL-RDASGATFTWPPSSGK--SAVGPPPERDLC 167
OY 283 STYRVSVYLTVLHODMLNGKEYCKVSNKALPAPIEKTISXAKGQPREPOVYTLPPSRDE 342
DB 168 GCYSISSVLPFGCAQPMNHGERTCTAHPBELKTPITANITKS-GTFPPEVHLPLPPSEE 226
OY 343 LTKNQ-VSLTCLVKGFPYSDIAVEMESNGQ--PENNYKTPPVLD-SDG--SFLLYSKLT 396
DB 227 LALNELVLTCLARGFSPDVLVRLQSGQLPREKYLTMASRQSPSGTTTFAVTSILR 286
OY 397 VDKSNQOGNNVSCGVNHEALHNHTOKSLSPG-----LQDLETC 438
DB 287 VAAEDWKKGDTFSQVNGHEALPLAFTOKTIDRLACKPTHVNVSVMAEVDGTC 339

RESULT 50
HYCS_HETFR STANDARD; PRT; 438 AA.
ID HVCS_HETFR
AC P23087;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region, secreted form (clone 3050).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=88328985; PubMed=338109;
RA Kokubu F., Hinde K., Litman R., Shambloot M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.";
RL EMBL J. 7:1979-1988(1988).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC EMBL: X07781, CA30613.1; ALT_INIT.
DR PIR: S01853; HYRKS.
DR HSSP: P01842; 7FAB.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003597; Ig_c1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 4.
DR SMART: SM00407; IgC1; 2.
DR PROSITE: PS50835; IG_LIKE; 4.
DR PROSITE: PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing.
FT NON TER 1
FT DOMAIN 1 99 CH1.
FT DOMAIN 100 205 CH2.
FT DOMAIN 206 308 CH3.
FT DOMAIN 309 418 CH4.
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 438 AA; 47920 MM; 9BD37C3BB07D15B CRC64;

Query Match 12.2%; Score 294; DB 1; Length 438;
Best Local Similarity 27.1%; Pred. No. 1e-12;
Matches 115; Conservative 65; Mismatches 153; Indels 92; Gaps 21;

OY 53 WKNSNQKIIGNGSFLTKPSKLDNADRRSLMDGNEPLIKNLKIEDSDYIEV- 111
DB 41 WKDNEPITTG-----LKTYPVLNKKGYTQ-----QSLTTESEVSSKTY-CEVR 88
OY 112 -----EDQKEVOLLVPG--LTANSPTHLQSGSLTLT-- 142
DB 89 RGSVWIKELIPDCKGDVNHPTVLITQSSSEIISRATVLCSTIDH---PSITVSWL 145
OY 143 -----LSPGSSPSVOCRSRPRGN--IQGKTLVSQLELDQSGTWTCTVLONOKKVEF 195
DB 146 KDQGHMSEGFVTSPT--C-----GVNGTFSATSRITVPAEFMTVKVYTCOV--SHQGVQ 197
OY 196 KIDI-----VCPAPBPSCDKHTCPPELLGSPVFLPBPXKDTLMISRTPEVTCVVDV 251
DB 198 SRNITGQVPC-----SCN-----DVIYKLLPSEIOVL-LEATVTLTCV---V 237
OY 252 SHEDPEVKFMVYDGVENHNAKTKPREEOYNSTRVVSVLTVLHODMLNGKEYCKVSNK 311
DB 238 SNAPYGVNVM-----TQOKSLKSEIANQPBEDADSVISTYNISTQAMLSGAEFCVYVNHQ 294
OY 312 ALPAPIEKTISKAKQP-REPOV-YTLPPSRDELTKNQVSLTCLVKGFPYSDIAVEMESN 369
DB 295 DLPTPLRAASHKEEVKDLREPSVILSPADVSAGFSLTCLVARGFPFRIEIVKWTVN 354
OY 370 GQPEN--NYKTTTPPVLDSD--GSFPLYSKLTVDVSRWQGVNFGSVNHEALHNHTOKSL 426
DB 355 DKSVPNGNYKTEVMAENDNSYFIYSLISIAEEMWASGASVGVNHEALPLKINRTV 414
OY 427 SLSPG 431
DB 415 NKSSG 419

RESULT 51
ALC_RABIT STANDARD; PRT; 299 AA.
ID ALC_RABIT
AC P01879;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

```

DE Ig alpha chain C region (Fragment).
OS Oryzology cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=84144059; PubMed=6322114;
RA Knight K.L., Martens C.L., Stoklosa C.M., Schneiderman R.D.;
RT "Genes encoding alpha-heavy chains of rabbit IgA: characterization of
RT cDNA encoding IgA-g subclass alpha-chains.";
RL Nucleic Acids Res. 12:1657-1670(1984).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- MISCELLANEOUS: This immunoglobulin belongs to the IgA-G subclass.
CC It was isolated from a rabbit homozygous FOR A2, N80, DE12, 15,
CC F71, G75 heavy chain haplotype.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X00353; CAA25100.1; -.
DR PIR; A02174; AHRB.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG LIKE; 2.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER 1 1
FT DOMAIN 1 167 IG-LIKE 1.
FT DOMAIN 174 276 IG-LIKE 2.
SQ SEQUENCE 299 AA; 32256 MW; 2512FE3F62E9A223 CRC64;
Query Match 12.1%; Score 293; DB 1; Length 299;
Best Local Similarity 31.7%; Pred. No. 7.6e-13;
Matches 78; Conservative 42; Mismatches 106; Indels 20; Gaps 10;
QY 200 VCPAPPEPSCDKHTHCP-----ELLG--PSVFLPPPKKDTLMSTRPEVTGVVNVDS 252
DB 41 LVPFPDCCPANSCTCTPSSSRNLISGQPSLISLRPLG-LLLGRASLTCTLSGK 99
QY 253 HEDPEKFMVYDGVENHAKTPREQYNSYRVSVTLVHODLNKEYKCKVSNKA 312
DB 100 NEDAV-FTW--EPTNGNEPVQORARDISGCTSVSVLPSSAEYTKATETFTCTVTHE 156
QY 313 LPA-PIEKTISKAKGQPREPVYTLPPSRDELTKN-QVSLTCLVKGFYPSDIAVWESNG 370
DB 157 IMSGTATISRGVTP--PQVHLPPPEELALNQEVTLTCLVGRGFSKDVLSWRHG 214
QY 371 Q--PENNY---KTPPVLDSDGFFLYSLTVDKSKWQGNVFGSGVMEALHNHTQKS 425
DB 215 QEVPEDSFLVWKSMPESSODKATYALTSLRVPADMDGDTYSQVGHGLAEHFTQKT 274
QY 426 LSLSPG 431
DB 275 IDRLAG 280

AC P23084;
DT 01-NOV-1991 (Rel. 20, Last Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain C region (Clone 6125) (Fragment).
OC Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88328985; PubMed=3138109;
RA Kohku F., Hinde K., Litman R., Shambloot M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
RT constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X07783; CAA30616.1; -.
DR PIR; S01851; HVRK05.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN 1 134 IG-LIKE 1.
FT DOMAIN 145 237 IG-LIKE 2.
FT DOMAIN 247 347 IG-LIKE 3.
FT CARBOHYD 98 98 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 132 132 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 343 343 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 347 347 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAc . . .) (POTENTIAL).
SQ SEQUENCE 370 AA; 40586 MW; 32833A6FE7ADE0 CRC64;
Query Match 11.9%; Score 286.5; DB 1; Length 370;
Best Local Similarity 26.2%; Pred. No. 2.7e-12;
Matches 101; Conservative 59; Mismatches 155; Indels 71; Gaps 16;
QY 72 GPKLNDADSRRSW-----DGNFLLIKNLIEDSD-----TYCEVDQKEE 117
DB 11 GSKTIYCEVRGSESMIWEIPDCKGDIVHPTVLTISSEETTSRFAVVCSTIIDFPE 70
QY 118 VOLVFGLTAN--SPTHLLGQSLTLTLSPGSSPVQCSRPRKNTIGGKTLSVSOLE 175
DB 71 -----ALTVMWLKDGHMESGFV-----SPAEITNG-----NFSATSRLLVPARE 111
QY 176 LODSGTWTCTVLQNKQKVEFKIDVPCAPAPPKSCDKHTCTPELLGSGVFLPPPKPDT 235
DB 112 WFTNTVYTCQVSHQATQSRNI-----TSPSPSECNH-----PAIKLLPSISIQV 157
QY 236 LMISRPETVCVVNDVSHEDPEVKMNVYDGVENHAKTPREBY-----NSTRVVSV 290
DB 158 L-LEATVTLTCV---VSNAPYGVNVSW-----TQEQKPLKSIIVAPGSDPDST 205
QY 291 LTVLHODWLNGKEYKCKVSNKALPAPIEKTISK-AKGQPREPVYTLPPSRDELTKN-V 348
DB 206 VDISTQAWLSEAVFVCVSHQDLPPLRDSIHKEMWKLRRPESVLLPRAPEISAEKFL 265

```

OY 349 SLTCLVKGFPSPDIKAVEMESNGOPEN--NYKTTPEVLUSD--GSFPLYSKLTVDKSRMQQG 405
DB 266 SLTCLVKGFPSPREIIVKNTVANDKSVNPGNYKTEVMAENDKSPFTISLSIAAEWASG 325
OY 406 NVFSCSVNHEALHNHYTOKSLSPG 431
DB 326 ASYSCVGHHEALPLKTIINRTVYKSSG 351

RESULT 53
HVC3_HETFR STANDARD; PRT; 393 AA.
ID HVC3_HETFR STANDARD; PRT; 393 AA.
AC P23086;
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
ID heavy chain C region (Clone 6121) (Fragment).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
  constant region genes in a phylogenetically primitive vertebrate.";
RL EMBL J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X07782; CAA30615.1; -.
CC PIR: S01852; HVRK1.
CC HSSP: P01842; 7FAB.
CC InterPro: IPR007110; Ig-like.
CC InterPro: IPR003597; Ig_c1.
CC InterPro: IPR003006; Ig_MHC.
CC Pfam: PF00047; Ig1; 3.
CC SMART: SM00407; Ig1; 2.
CC PROSITE: PS00835; IG_LIKE; 3.
CC PROSITE: PS00290; IG_MHC; 3.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 1 157
FT DOMAIN 2 168
FT DOMAIN 3 270
FT CAROHD 119
FT CAROHD 155
FT CAROHD 200
FT CAROHD 230
FT CAROHD 329
FT CAROHD 366
FT CAROHD 370
FT CAROHD 380
FT CAROHD 393
FT SEQUENCE 393 AA; 43081 MW; 4844076972F18B5 CRC64;

Query Match 11.6%; Score 281; DB 1; Length 393;
Best Local Similarity 26.4%; Pred. No. 6,9e-12;
Matches 115; Conservative 69; Mismatches 161; Indels 90; Gaps 21;
OY 23 TQGNKV---VLAKGK---DTVELTCTASOKSISQFHMKNNOIKILNGSFLTYGPKSL 76
DB 4 TTGLKTVPSVINKKGTYSQSLITTESB-----VGSSKI 38

```

```

OY 77 NDRADSRRLMDGNPFLIKNLKIEDSDTYICEVEDQKEEVOLVFG--LTANSDTHL 134
DB 39 YCEVRREESVWIKK-----IPDCKGDKVHPFTVILITSSSEIISRRATVLCISIDH-- 91
OY 135 QGQSLTLT-----LESPPGSSPSVQCRSPRGKN--IOGKTLVSQLELDSDGTWCT 185
DB 92 -PESITVSWLKDQGHMSEGFVTSPT--C-----GVNGFSAISRLLTVARBEFTMKVYTCQ 144
OY 186 VLNQKQKVEFKIDI---VCPAPBPKSCDKTHTCPBLLGSPVFLFPPEPKDTLMISRT 241
DB 145 V--SHQGVTSRNTGSQVPC-----SCN-----DPVTKLPLPSIEQVL-LEAT 185
OY 242 PEVTCVVVDVSHEDPEVKFMNVDGVEVHNAKTRPREQYNSTYRVYSLTVLHQDWLNG 301
DB 186 VTLNCT---VSNAPYGVNSW---TOEQKSLKSEIAVQPEBDADSVISTYNIQTAMLSG 239
OY 302 KEYKCKVSNALPAPKEKTSKAKGP-REPOV-YTLPPSRDELTKNQVSLTCLVKGFYP 359
DB 240 AEFYCVVNHQDLPPLRASLHKKEVVDLRPEPSVILSPADVSAGRFSLTCLVRGFSF 299
OY 360 SDIAYEVESNGOPEN--NYKTTPEVLUSD--GSFPLYSKLTVDKSRMQGAVFSCVNHAE 416
DB 300 RELFVKWTINDKSVNPGNYKTEVMAENDNSYFIYLSIAAEWASGASYSCVGHAE 359
OY 417 LHNHYTOKSLSPG 431
DB 360 IPLKTIINRTVYKSSG 374

RESULT 54
ALC1_GORGO STANDARD; PRT; 353 AA.
ID ALC1_GORGO STANDARD; PRT; 353 AA.
AC P20758;
DT 01-FEB-1991 (Rel. 17, Created)
DE 01-FEB-1991 (Rel. 17, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig alpha-1 chain C region.
GN IGHAI1
OS Gorilla gorilla gorilla (lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph node;
RX MEDLINE=89386006; PubMed=2506527;
RA Kawamura S., Omoto K., Ueda S.;
RT "Nucleotide sequence of the gorilla immunoglobulin alpha 1 gene.";
RL Nucleic Acids Res. 17:6732-6732(1989).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
  secretions. It may serve both to defend against local infection
  and to prevent access of foreign antigens to the general
  immunologic system.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X15045; CAA3147.1; ALT_INT.
CC HSSP: P01810; 2FBJ.
CC InterPro: IPR007110; Ig-like.
CC InterPro: IPR003597; Ig_c1.
CC InterPro: IPR003006; Ig_MHC.
CC Pfam: PF00047; Ig1; 3.
CC SMART: SM00407; Ig1; 2.
CC PROSITE: PS00835; IG_LIKE; 3.
CC PROSITE: PS00290; IG_MHC; 1.

```

```

KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON TER 1 1
FT DOMAIN 6 98
FT DOMAIN 125 220
FT DOMAIN 228 330
FT DISULFID 14 14
FT DISULFID 26 85
FT DISULFID 77 101
FT DISULFID 122 122
FT DISULFID 123 180
FT DISULFID 147 204
FT DISULFID 182 182
FT DISULFID 192 192
FT DISULFID 250 313
FT DISULFID 352 352
FT CARBOHYD 144 144
FT CARBOHYD 340 340
SQ SEQUENCE 353 AA, 37755 MW, 48208DB02ACT514 CRC64;

Query Match
Best Local Similarity 28.0%; Score 279; DB 1; Length 353;
Matches 96; Conservative 41; Mismatches 126; Indels 80; Gaps 16;

QY 127 ANSDTHLQGGSLTLTLESPPSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTV 186
DB 59 ASGDYLTSSQ---LTL-----PATQC--PDGKSV-----TCHV 87
QY 187 LQNGKVKFEKIDIVC-----PAPEKSCDKHTTCPELGGSPVLPFPKP 232
DB 88 --NHNTNSQDVTPQKRVSTPTPTSPSTPPSPPCCH-----PFLSLHRL 134
QY 233 KDTLMSIRPEVTCVVDVSHEDPEVKFMVYDGVENHAKKPRBEQYNSTRVSVLT 292
DB 135 ED-LILGSEANLTCLTGL-RDASGVTFTWTPSSGK--SAGVGPERRDLGCVSVSVLP 190
QY 293 VLIHQWLNGKEYKCKVSNKALPAPIEKITSKAKGQPREPOVYTLPPSRDELTKNQ-VSLT 351
DB 191 GCAEPMNHGKFTCTCAAYAPESKTPLATLSKS-GMMFRPEVHLLPPSEHLLNLVLT 249
QY 352 CLVKEPYPSDIAVEMESNQ--PENNYKTPPVLD-SG--SFPLYSLTLVTKSRMOQN 406
DB 250 CLARGSPDVLVLRWLGSSQELPREKXYLTWASRQEPSQGTTFPAVTSILRVAEDMKGD 309
QY 407 VFSCGVMEALHNHYTKSLSPG-----LQDFTC 438
DB 310 TFSCHVGHALLPLAFTQKTIIDLAKGPHVNVSVMAAEVDGTC 352

RESULT 55
ALCI HUMAN STANDARD; PRT; 353 AA.
ID _ALCI_HUMAN
AC P01876;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG alpha-1 chain C region.
GN IGHAI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84130179; PubMed=6421489;
RA Planagan J.G., Lefranc M.-P., Rabbits T.H.;
RT "Mechanisms of divergence and convergence of the human immunoglobulin
RL alpha 1 and alpha 2 constant region gene sequences.";
RN Cell 36:681-688(1984).
RN [2]

```

```

RP SEQUENCE (MYELOMA PROTEIN BUR), AND DISULFIDE BONDS.
RX MEDLINE=79151016; PubMed=107164;
RA Putnam F.W., Liu Y.-S.V., Low T.L.K.;
RT "Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal
RT IgA1 protease, digestion, Fab and Fc fragments, and the complete
RT amino acid sequence of the alpha 1 heavy chain.";
RL J. Biol. Chem. 254:2865-2874(1979).
RN [3]
RP SEQUENCE (MYELOMA PROTEIN TRO).
RX MEDLINE=76023781; PubMed=809331;
RA Kratzin H., Altevogt P., Ruban E., Kort A., Staroscik K.,
RA Hilschmann N.;
RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.),
RT II. The amino acid sequence of the H-chain, alpha-type, subgroup II;
RT structure of the complete IgA-molecule.";
RL Hope-Seyler S Z. Physiol. Chem. 356:1337-1342(1975).
RN [4]
RP DISULFIDE BONDS.
RX MEDLINE=80114124; PubMed=393607;
RA Yang C.-Y., Kratzin H., Gotz H., Hilschmann N.;
RT "Rule of antibody structure. Primary structure of a human monoclonal
RT IgA-immunoglobulin (myeloma protein Tro). VII. Purification and
RT characterization of the disulfide bridges.";
RL Hope-Seyler S Z. Physiol. Chem. 360:1919-1940(1979).
RN [5]
RP REVIEW.
RX MEDLINE=91054387; PubMed=2241915;
RA Kerr M.A.;
RT "The structure and function of human IgA.";
RL Biochem. J. 271:285-296(1990).
CC -I- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -I- SUBUNIT: Monomeric or polymeric.
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; J00220; AAC82528.1; ALI_INIT.
DR PIR; A23360; AIHU.
DR PDB; 1IGA; 15-JUN-99.
DR Genew; HGNC:5478; IGHAI.
DR MIM; 146900;
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_c1.
DR SMART; SM00407; Igc1; 2.
DR PROSITE; PS50835; IG-LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Polymorphism; 3D-structure; Repeat.
FT NON TER 1 1
FT DOMAIN 6 98
FT DOMAIN 125 220
FT DOMAIN 228 330
FT DISULFID 14 14
FT DISULFID 26 85
FT DISULFID 77 101
FT DISULFID 122 122
FT DISULFID 123 180
FT DISULFID 147 204
FT DISULFID 182 182
INTERCHAIN (WITH HEAVY CHAIN) (OR 180, IN
REF. 4).

```

```

FT DISULFID 192 192 INTERCHAIN (WITH HEAVY CHAIN OF ANOTHER
FT SUBUNIT) (PROBABLE).
FT DISULFID 250 313 INTERCHAIN (WITH J CHAIN).
FT DISULFID 352 105 O-LINKED.
FT CARBOHYD 105 105 O-LINKED.
FT CARBOHYD 111 111 O-LINKED.
FT CARBOHYD 113 113 O-LINKED.
FT CARBOHYD 119 119 O-LINKED.
FT CARBOHYD 121 121 O-LINKED.
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .).
FT VARIANT 176 176 E-> D (in dSNP:1407).
FT CONFLICT 163 165 TPS-> PST (IN REF. 2).
FT CONFLICT 176 176 E-> B (IN REF. 3).
FT CONFLICT 190 190 P-> S (IN REF. 3).
FT CONFLICT 227 227 R-> H (IN REF. 3).
FT CONFLICT 231 231 H-> R (IN REF. 3).
FT CONFLICT 290 290 T-> E (IN REF. 3).
SQ SEQUENCE 353 AA; 37654 MW; EBA1ECB7E85DB21 CRC64;

```

```

Query Match 11.5%; Score 277.5; DB 1; Length 353;
Best Local Similarity 27.2%; Pred. No. 1e-11;
Matches 104; Conservative 51; Mismatches 153; Indels 75; Gaps 18;

```

```

QY 92 FPLIKNKIEDSDTYICEVEDQKEEVQLVF-----GLT-----ANSDTHLLQG 136
DB 9 FPLISCTSPQDPGNVIACTVGCFPPQEPPLSTWSSSGGVTAARNPPSQDASGLDYTTSS 68
QY 137 QSLTTLTSPSSSPSSVQCRSPRGKNIQGGKTLVSQLELDGSGTCTCTVLONQKVEFK 196
DB 69 Q-----LTL-----PATQCLA--GKSV-----TCHVHYNTNPSQ-- 95
QY 197 IDI--VPCAPAPKSCDKHTHCPELLG---PSVFLFPPKPKDTLMISTPEVTCVVVDVS 252
DB 96 -DVTPPCVSPSTPPSPSPSTPPSPSCCHPLSLHRALEDL-LTLGSEANLTCITLGL- 152
QY 253 HEDPEVKENWYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHODMLNKEKCKVSNKA 312
DB 153 RDASVTFTWTPSSSK--SAVGPPERDLGCGYSVSLPGCAEPMNNGKFTCTAAYPE 210
QY 313 LPAPEKTIKAKGQPREPOVYTLTPSRDELTKNQ-VSLTCLVKGFPYSDIAVESNGQ 371
DB 211 SKTPLTALISKS-GVTFRPEVHLFPPSEELALNLELVLTCLAGFSKDYLVWMLQSQ 269
QY 312 --PENNYTTPPYLD-SDG--SFPLYSLTVDKSNQGNFSCSVMEALHNHYTKSL 426
DB 270 ELPRKXYLTWASROEPPSGTTFPAVTSLRVAAEDMKKGDTFSCVNGHEALPLAFTQKTI 329
QY 427 SLSPG-----LQIDETC 438
DB 330 DRLAGKPTHVNVVMAEVDGTC 352

```

```

RESULT 56
MUCM ICTPU STANDARD; PRT; 481 AA.
AC P23735:
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig mu chain C region membrane-bound form.
OS Ictalurus punctatus (Channell catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384824; PubMed=2119496;
RA Wilson M.R., Marcuz A., van Ginkel F., Miller N.W., Clem L.W.,
RA Middleton D., Watt G.W.;
"the immunoglobulin M heavy chain constant region gene of the channel

```

```

RT catfish, Ictalurus punctatus: an unusual mRNA splice pattern produces
RT the membrane form of the molecule."
RL Nucleic Acids Res. 18:5227-5233(1990).
CC -I- MISCELLANEOUS: During differentiation, B lymphocytes switch from
CC expression of membrane-bound IgM to secretion of IgM. The mu
CC chains of membrane and secreted IgM differ in their C-terminal
CC segments.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X52617; -; NOT_ANNOTATED_CDS.
DR HSSP; P01857; IFC1.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane.
KW NON TER 1 1
FT DOMAIN 1 101 CH1.
FT DOMAIN 102 203 CH2.
FT DOMAIN 204 304 CH3.
FT DOMAIN 305 434 CH4.
FT TRANSMEM 459 478 POTENTIAL.
FT DOMAIN 479 481 CYTOPLASMIC.
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN)
FT DISULFID 14 14 (BY SIMILARITY).
FT DISULFID 26 85 (BY SIMILARITY).
FT DISULFID 129 184 BY SIMILARITY.
FT DISULFID 233 284 BY SIMILARITY.
FT DISULFID 329 391 BY SIMILARITY.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 481 AA; 53337 MW; 668E784E181C71BA CRC64;

```

```

Query Match 11.4%; Score 275; DB 1; Length 481;
Best Local Similarity 23.4%; Pred. No. 2.2e-11;
Matches 108; Conservative 68; Mismatches 145; Indels 140; Gaps 22;

```

```

QY 37 VELTCTASQKS---IOFHKNSNQIKILGNQSFLLTKGSKLNDRADSRSLMDQNP 93
DB 22 VTLGCVTRDLASADGLSFIWKDAS-----GSALT-----DVQYP 56
QY 94 LI-----IKNLIKEDSD-----TYICEVE-----DQKEVOLVGLT 126
DB 57 AVQATGTYTSVSHRYVASDMNNGKFTCEVKNGLSKDASLQKPYREILHASTL---LT 113
QY 127 ANSDTHLLQGSQSLTLTSPSSSPVQCRSPRG-----KNY---QGG 166
DB 114 TPQTEIDNGTAFVCLATP-----FSPSHFPKWTLEKTDISNKKYKENIVSQNGK 164
QY 167 KTLASVSGLELDGSGTW-----CYLONQKQYEFKIDIVPCAPAPKSCDKHTHCPEL 219
DB 165 NFTAISVLELSAS-EMWSTSTSPVKCFEQKNNHVFKEASVAPDXTQPO-----VK 214
QY 220 LGGSPVFLPFPKPKDITMISTPEVTCVVVDVSHEDPEVK-FMVYVDGVEVNAKTKPRE 278
DB 215 ITGPST-----EDILIRAGQLEC---RAEGDGFKSIKMLINREISSLSNL--- 259
QY 279 EGVNSTYRVSVLT-VLHODMLNKEKCKVSNKA LPAPEK-TISKAKGQPREPOVYTL 336

```

```

Db      260 -----SSKTVSLQTHIGFEEMWINGTEFFICEVHEAFTQGVKATFRKENGNEPFXVYL 315
Qy      337 PSSRDELTKNOVSLTCLVGFPSDIAVWESN-----GQENNYKTPPVLDSDGS 388
Db      316 AP--PSSGESSTLTCTCYVDFFPKAVAVSWLVNDKQVEEVEYEDN---TVAVIDRNL 369
Qy      369 FFLYSKLTVDKSRMOGCVFSCSVMEALHNHYTKSLIS 429
Db      370 FSVISOLIKTKADMNSGVFSCLVTHESIKDCVRHISRSL 410

RESULT 57
ALC_MOUSE STANDARD; PRT; 344 AA.
AC P01878;
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 15-MAR-2004 (rel. 43, Last annotation update)
DE Ig alpha chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (MYELOMA ABB48 AND J558).
RA Aultfray C., Negeotte R., Sikorav J.-L., Heidmann O., Rougeon F.;
RT "Mouse immunoglobulin A: nucleotide sequence of the structural gene
RL for the alpha heavy chain derived from cloned cDNAs."
RN [2]
RP SEQUENCE OF 1-213 (MOPC 47A).
RA MEDLINE=80049769; PubMed=115869;
RA Robinson E.A., Appella E.;
RT "Amino acid sequence of a mouse myeloma immunoglobulin heavy chain
RL (MOPC 47 A) with a 100-residue deletion."
RN [3]
RP SEQUENCE OF 1-254 AND 291-344 (M511).
RA MEDLINE=81054880; PubMed=6776528;
RA Robinson E.A., Appella E.;
RT "Complete amino acid sequence of a mouse immunoglobulin alpha chain
RL (MOPC 511)."
RN [4]
RP Proc. Natl. Acad. Sci. U.S.A. 77:4909-4913(1980).
RT "FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- MISCELLANEOUS: The final C-region domain is deleted from Ref.2
CC chain. It was isolated from a myeloma protein that contains 1
CC light and 1 heavy chain per molecule, linked by a disulfide bond.
CC In contrast, normal mouse IgA molecules contain 2 light and 2
CC heavy chains and lack a light-heavy chain disulfide bond.
CC -1- MISCELLANEOUS: M511 chain was isolated from a myeloma protein that
CC binds phosphorylcholine.
CC -1- MISCELLANEOUS: M511 sequence was compared with that of mouse MOPC
CC 47A, and a genetic mechanism for the deletion of the CH3 domain of
CC the mutant chain is proposed.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D11468; BAA02026.1; ALT_INITT.
CC PIR; A91479; AHMS.
CC HSSP; P01810; 2FBJ.
CC InterPro; IPR007110; Ig-like.

```

```

DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF000477; Ig_3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.

FT DOMAIN 1 6 99 IG-LIKE 1.
FT DOMAIN 116 206 IG-LIKE 2.
FT DOMAIN 219 321 IG-LIKE 3.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 76 100 BY SIMILARITY.
FT DISULFID 114 171 BY SIMILARITY.
FT DISULFID 138 195 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .).
FT CARBOHYD 101 101 O-LINKED.
FT CARBOHYD 331 331 O-LINKED.
FT VARIANT 18 2 S -> A (IN MOPC 47A).
FT VARIANT 67 18 S -> C (IN MOPC 47A AND M511).
FT VARIANT 73 73 N -> S (IN MOPC 47A AND M511).
FT VARIANT 112 112 A -> T (IN MOPC 47A).
FT VARIANT 135 135 P -> G (IN M511).
FT VARIANT 141 141 S -> Q (IN MOPC 47A AND M511).
FT VARIANT 168 168 N -> D (IN MOPC 47A AND M511).
FT VARIANT 212 213 Q -> E (IN MOPC 47A).
FT VARIANT 235 235 VT -> SO (IN MOPC 47A).
FT VARIANT 255 235 B -> G (IN M511).
FT VARIANT 295 290 MISSING (IN M511).
FT VARIANT 301 301 T -> D (IN M511).
FT VARIANT 301 301 Q -> G (IN M511).
FT VARIANT 329 329 N -> Q (IN M511).
FT VARIANT 331 331 S -> N (IN M511).
SQ SEQUENCE 344 AA; 36876 MW; 3694CF9B19A9F8 CRC64;

Query Match 11.0%; Score 264.5; DB 1; Length 344;
Best Local Similarity 26.7%; Pred. No. 7,5e-11;
Matches 85; Conservative 57; Mismatches 117; Indels 59; Gaps 15;

Qy 136 GQSLTTLTLESPPG-----SSPVQCRSPKKNITQGGKTLVSQLELDQSGTW 182
Db 45 GKDIR-TYNFPPLALASGRYTMSNQLTLPVAVCC--PEGESYK----- 83
Qy 183 TCTVQNGKQKVEFKIDIVPCAPPEKSCDKHTGCELLGSPVFLFPKPKDTMISRTP 242
Db 84 -CSVQSDSNPQV-ELVD-VNCSGPTPPPTITPSCQ----PSLSIQRLALVD-LTLGSDA 134
Qy 243 EYTCVVVDSHEDPEVKENMYVDGVEVNAKTKPREQYNST---YRVVSVLTVLHQDML 299
Db 135 SITCTLNLGRNPEQAV-FTW-----EPSTGDAVQKAVQNSCGYSVSVLPQCAERNN 188
Qy 300 NGKEYKCVSNKALPAPIEKTIKSKAGQRPBPQVYTLPPSDELTKNO-VSLTCLVKGFPY 358
Db 189 SGASPKCTVTHPE-SGTLTGTLAKYTVNTFPQVHLPLPPEBELNLNLSTLCVRAFNN 247
Qy 359 PSDIAVWESNGQ---PENNYKTPPVLDSDG--SFFLYSKLTVDKSRMOGCVFSCSVM 413
Db 248 PREVLVRLHGNBELSPSYLVFEPFLKEPBGATTVLTVSLVRAETWKGQDQYSCWVG 307
Qy 414 HEALHNHYTKSLISLSPG 431
Db 308 HEALPMNFTQKTIIDRLSG 325

RESULT 58
SHS1_MOUSE STANDARD; PRT; 513 AA.
AC P97757; O08907; O35924; O88555; O88556; P97796; Q8R559; Q9QX57;
AC Q9WTN4;
DT 10-OCT-2003 (rel. 42, Created)
DT 10-OCT-2003 (rel. 42, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor

```

DE (SHP substrate-1) (SHPS-1) (inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (SIRP-alpha-1) (msIRP-alpha1) (MyD-1
DE antigen) (Brain Ig-like molecule with tyrosine-based activation
DE motif) (Bit) (p84).
CN PTPNS1 OR SHPS1 OR SIRP OR MYD1 OR BIT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=97223399; PubMed=9070220;
RA Yamao T., Matozaki T., Amano K., Matsuda Y., Takahashi N., Ochi F.,
RA Fujioka Y., Kasuga M.;
RT "Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal
RT localization of genes."; Biochem. Biophys. Res. Commun. 231:61-67(1997).
RL Biochem. Biophys. Res. Commun. 231:61-67(1997).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND VARIANTS ALA-29; ARG-67; ARG-91;
RP THR-96; SER-128; PRO-194 AND ASN-224.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=97230468; PubMed=9073522;
RA Ohnishi H., Kubota M., Sano S.-I.;
RT "Bit (Bit) maps to mouse chromosome 2."; Genomics 40:504-506(1997).
RL Genomics 40:504-506(1997).
[3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), SEQUENCE OF 32-53 AND 422-433
RP (ISOFORM 2), VARIANTS ALA-29; ARG-91; THR-96 AND SER-128,
RP N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RC STRAIN=BALB/c; TISSUE=Brain, and Cerebellum;
RX MEDLINE=98012243; PubMed=9348339;
RA Comu S., Weng W., Olinkev S., Ishwad P., Mi Z., Hempel J., Watkins S.,
RA Lagenauer C.F., Narayanan V.;
RT "The murine P84 neural adhesion molecule is SHPS-1, a member of the
RT phosphatase-binding protein family."; J. Neurosci. 17:8702-8710(1997).
RL J. Neurosci. 17:8702-8710(1997).
[4]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365,
RP N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, INTERACTION
RP WITH PTPN6, AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6; TISSUE=Fetal thymus;
RX MEDLINE=98380500; PubMed=9712903;
RA Veilleux A., Thibaudan E., Latour S.;
RT "High expression of inhibitory receptor SHPS-1 and its association
RT with protein tyrosine phosphatase SHP-1 in macrophages."; J. Biol. Chem. 273:22719-22728(1998).
RL J. Biol. Chem. 273:22719-22728(1998).
[5]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365, AND
RP N-GLYCOSYLATION.
RC STRAIN=129/SV, and C57BL/6; TISSUE=Brain, and Liver;
RX MEDLINE=20053880; PubMed=10585853;
RA Sano S.-I., Ohnishi H., Kubota M.;
RT "Gene structure of mouse BIT/SHPS-1."; RT Biochem. J. 344:667-675(1999).
RL Biochem. J. 344:667-675(1999).
[6]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365.
RC STRAIN=ILS, and ISS;
RX MEDLINE=21363810; PubMed=11471062;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Ganniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikele J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related QTLs."; Mamm. Genome 12:657-663(2001).
RL Mamm. Genome 12:657-663(2001).
[7]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT PRO-194.
RC TISSUE=Placenta;

RA Wang H., Chen Z., Ullrich A.;
RT "Epidermal growth factor-induced association of SHP2 with mouse SIRP-
RT alpha1."; RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
[8]
RP FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=90152134; PubMed=2303162;
RA Chuang W., Lagenauer C.F.;
RT "Central nervous system antigen P84 can serve as a substrate for
RT neurite outgrowth."; Dev. Biol. 137:219-232(1990).
RL Dev. Biol. 137:219-232(1990).
[9]
RP N-GLYCOSYLATION, PHOSPHORYLATION BY JAK2 IN RESPONSE TO GROWTH
RP HORMONE, AND INTERACTIONS WITH JAK2 AND PTPN11.
RX MEDLINE=98175985; PubMed=907023;
RA Stofega M.R., Wang H., Ullrich A., Carter-Su C.;
RT "Growth hormone regulation of SIRP and SHP-2 tyrosyl phosphorylation
RT and association."; J. Biol. Chem. 273:7112-7117(1998).
RL J. Biol. Chem. 273:7112-7117(1998).
[10]
RP INTERACTION WITH CD47, AND TISSUE SPECIFICITY.
RX MEDLINE=99091586; PubMed=9872987;
RA Jiang P., Lagenauer C.F., Narayanan V.;
RT "Integrin-associated protein is a ligand for the P84 neural adhesion
RT molecule."; J. Biol. Chem. 274:559-562(1999).
RL J. Biol. Chem. 274:559-562(1999).
[11]
RP INTERACTION WITH FGR.
RX MEDLINE=20130295; PubMed=10662797;
RA Gresham H.D., Dale B.M., Potter J.W., Chang P.W., Vines C.M.,
RA Lowell C.A., Lagenauer C.F., Willman C.L.;
RT "Negative regulation of phagocytosis in murine macrophages by the Src
RT kinase family member, Fgr."; J. Exp. Med. 191:515-528(2000).
RL J. Exp. Med. 191:515-528(2000).
[12]
RP FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
RP as docking protein and induces translocation of PTPN6,
RP PTPN11 and other binding partners from the cytosol to the
RP plasma membrane. Supports adhesion of cerebellar neurons, neurite
RP outgrowth and glial cell attachment. May play a key role in
RP intracellular signaling during synaptogenesis and in synaptic
RP function. Involved in the negative regulation of receptor tyrosine
RP kinase-coupled cellular responses induced by cell adhesion, growth
RP factors or insulin. Mediates negative regulation of phagocytosis,
RP mast cell activation and dendritic cell activation. CD47 binding
RP prevents maturation of immature dendritic cells and inhibits
RP cytokine production by mature dendritic cells (By similarity).
CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
CC macrophages, where it primarily binds PTPN6. Binds GRB2 vitro.
CC Binds FGR. Binds JAK2 irrespective of its phosphorylation status
CC and forms a stable complex. Binds SCAP1 and/or SCAP2. The
CC resulting complex recruits FVB. Binds PTK2B (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1; Synonyms=a1;
CC IsoId=P97797-1; Sequence=Displayed;
CC Name=2; Synonyms=a', Large;
CC IsoId=P97797-2; Sequence=VSP_007032;
CC Name=3; Synonyms=b, Small;
CC IsoId=P97797-3; Sequence=VSP_007031;
CC -1- TISSUE SPECIFICITY: Highly expressed in cerebral cortex, brain,
CC spinal cord, cerebellum and spleen, and at much lower levels in
CC kidney, thymus, heart, lung and liver. Within the cerebellum,
CC highly expressed throughout the molecular layer, and in synaptic
CC glomeruli in the granule cell layer. Detected in neurons of the
CC hippocampus and dentate gyrus, and in olfactory bulb. Not detected
CC in Purkinje cells. Highly expressed in the plexiform layers, optic
CC fiber layer and the outer segments of the photoreceptor layer in
CC the retina. Highly expressed in macrophages. Isoform 3 is detected
CC at very low levels in all tissues tested.
CC -1- DEVELOPMENTAL STAGE: Highly expressed in the CNS of embryos from
CC day 7 to 17.

```
CC -1- PTM: N-glycosylated.
CC -1- PTM: Phosphorylated on tyrosine residues.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D87967; BAA13520.1; -
DR EMBL; D87968; BAA13521.1; -
DR EMBL; D87985; BAA20376.1; -
DR EMBL; U89694; AAB92591.1; -
DR EMBL; AF072543; AAC24886.1; -
DR EMBL; AF072544; AAC24887.1; -
DR EMBL; AB024507; BAA89290.1; -
DR EMBL; AB024508; BAA89290.1; JOINED.
DR EMBL; AB024501; BAA89290.1; JOINED.
DR EMBL; AB024502; BAA89290.1; JOINED.
DR EMBL; AB024503; BAA89290.1; JOINED.
DR EMBL; AB024504; BAA89290.1; JOINED.
DR EMBL; AB024505; BAA89290.1; JOINED.
DR EMBL; AB024506; BAA89290.1; JOINED.
DR EMBL; AB018194; BAA76555.1; -
DR EMBL; AB024507; BAA89289.1; -
DR EMBL; AB024508; BAA89289.1; JOINED.
DR EMBL; AB024501; BAA89289.1; JOINED.
DR EMBL; AB024502; BAA89289.1; JOINED.
DR EMBL; AB024503; BAA89289.1; JOINED.
DR EMBL; AB024504; BAA89289.1; JOINED.
DR EMBL; AB024505; BAA89289.1; JOINED.
DR EMBL; AB024506; BAA89289.1; JOINED.
DR EMBL; AF332079; AAK56107.1; -
DR EMBL; AF332080; AAK56108.1; -
DR EMBL; Y10349; CAA71375.1; -
DR PIR; JCS289; JCS289.
DR MGD; MGI:108563; Pcpn1.
DR GO; GO:000580; F.cytoskeletal regulator activity; IMP.
DR GO; GO:0045309; F.phosphoprotein amino acid binding; IPI.
DR GO; GO:0005515; F.protein binding; IMP.
DR GO; GO:0007015; F.actin filament organization; IMP.
DR GO; GO:0006928; P.cell motility; IMP.
DR GO; GO:0007160; P.cell-matrix adhesion; IMP.
-----
Query Match 8.2%; Score 198; DB 1; Length 513;
Best Local Similarity 21.3%; Pred. No. 3.7e-06;
Matches 93; Conservative 57; Mismatches 151; Indels 136; Gaps 17;
-----
QY 23 TGSNKVVLKGGDTVELTCTASQKSIQPHMKNKNOIKILGNGSFLTKGPKSLNDRADS 82
DB 37 TQPEKSVSAAGSDTVLNCITLT-----SLTPVGPPIK----- 67
QY 83 RSLNMQGFPRLIKLKIEDSDTYICEVEDQKEVQLVFGITANSDFHLLOGOSLTIT 142
DB 68 ---WYRG-----VGGSRLLIYSPFG---HF----- 87
QY 143 LESPSPSSPVQCRSPRGKNIQGGKTLVSQLELOSGTWTCTVLONQKVEPKIDIVPC 202
DB 88 -----PRVTNVSATKRNKMDIFIRISNTPEDAGTYCYKFG----- 125
QY 203 PAPEPKSCDKHTCPPELLAGPSVFLF---PKDTLMI SRTPEVTCVVVDVSH-- 253
DB 126 -----KGSPEDTEIQSGGTEVYLAKSPPEVSGPADRGIPDQKVNFTC---KSHGF 176
QY 254 EDEPEVFNMYVGVCEVHNKATKREQVNSTYKRVSVLVTLVLMODMLNGEKYCKKSNKLL 313
DB 177 SPRLNLTLMKFKQGEHLHLETTVTPSGKVNYSISTVAV---LNSMDVSHSKYICEVA 232
QY 314 PAPIEKTIKAKQGPPEPQVYTLPS-----RDELTKNQVSLTCLVKGFYPSDIAVEMWS 368
```

```
DB 223 HITLDR--SPLRGIANLSNFRIVSPVTKVQGSPPSPNQVNLTCABRRYPEDLQILWE 290
QY 369 NQPEPNKTKTPPVV--DSDSFELYSKLTUDKSNQOGANVSCGVME--ALNNHYT- 422
DB 291 NQNVARN--DPEKNTLKTKTDGTGYNTYSLFLVNSSAHREDVFTCOVKHDQDPAITRNHTV 348
QY 423 -----QKSLSPG 431
DB 349 LGLAHSDDGSMQTRPG 365
-----
RESULT 59
SHS1 BOVIN
ID SHS1 BOVIN STANDARD; PRT; 506 AA.
AC O46531; O46632;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor
DE (SHP substrate-1) (SHPS-1) (Inhibitory receptor SHS-1) (Signal-
DE regulatory protein alpha-1) (Stip-alpha-1) (Myd-1 antigen).
GN PRPS1 OR SHPS1 OR SIRP OR MYD1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OC NCB1_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS SER-23; ALA-28; LEU-61; ARG-70;
RP HIS-120; I25-GLN; I27-GLY; I29-HIS; I32-VAL; ASN-145; VAL-153;
RP ASP-203; ARG-261; LEU-302; LEU-316; ARG-337; ASN-367; LEU-422; PHE-429
RP AND GLU-433.
RC STRAIN=Frilesian; TISSUE=Peripheral blood;
RX MEDLINE=98143722; PubMed=9485180;
RA Brooke G.P., Parsons K.R., Howard C.J.;
RT "Cloning of two members of the SIRP alpha family of protein tyrosine
RT phosphatase binding proteins in cattle that are expressed on monocytes
RT and a subpopulation of dendritic cells and which mediate binding to
RT CD4 T cells."
RT Eur. J. Immunol. 28:1-11(1998).
RL
RL -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
CC as docking protein and induces translocation of PRPN6,
CC PRPN1 and other binding partners from the cytosol to the
CC plasma membrane. Supports adhesion of cerebellar neurons, neurite
CC outgrowth and glial cell attachment. May play a key role in
CC intracellular signaling during synaptogenesis and in synaptic
CC function. Involved in the negative regulation of
CC receptor tyrosine kinase-coupled cellular responses induced by
CC cell adhesion, growth factors or insulin. Mediates negative
CC regulation of phagocytosis; mast cell activation and dendritic
CC cell activation. CD47 binding prevents maturation of immature
CC dendritic cells and inhibits cytokine production by mature
CC dendritic cells (By similarity).
CC -1- SUBUNIT: Binds PRPN1 when tyrosine-phosphorylated, except in
CC macrophages, where it primarily binds PRPN. Binds GBB2 in vitro.
CC Binds JAK2 irrespective of its phosphorylation status and forms a
CC stable complex. Binds SCAP1 and/or SCAP2. The resulting complex
CC recruits Fyb. Binds FGR and PRK2B (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Highly expressed in spleen macrophages.
CC Detected in skin dendritic cells.
CC -1- PTM: Phosphorylated on tyrosine residues (By similarity).
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL; Y11045; CAA71942.1; -.
DR EMBL; Y11046; CAA71943.1; -.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Repeat; Signal; Transmembrane; Immunoglobulin domain; SH2-binding;
KW Glycoprotein; Phosphorylation; Polymorphism.
FT SIGNAL 1 29
FT CHAIN 30 506
FT DOMAIN 30 371
FT TRANSMEM 372 392
FT DOMAIN 393 506
FT DOMAIN 30 145
FT DOMAIN 148 248
FT DOMAIN 255 348
FT DISULFID 55 121
FT DISULFID 170 228
FT DISULFID 273 331
FT SITE 432 435
FT SITE 441 446
FT SITE 455 458
FT SITE 472 475
FT SITE 498 501
FT MOD_RES 431 431
FT MOD_RES 455 455
FT MOD_RES 472 472
FT MOD_RES 498 498
FT MOD_RES 92 92
FT CARBOHYD 167 179
FT CARBOHYD 204 204
FT CARBOHYD 210 210
FT CARBOHYD 246 246
FT CARBOHYD 270 270
FT CARBOHYD 292 292
FT CARBOHYD 311 311
FT CARBOHYD 319 319
FT CARBOHYD 344 344
FT CARBOHYD 365 365
FT CARBOHYD 368 368
FT VARIANT 23 23
FT VARIANT 28 28
FT VARIANT 61 61
FT VARIANT 70 70
FT VARIANT 120 120
FT VARIANT 125 125
FT VARIANT 127 127
FT VARIANT 129 129
FT VARIANT 132 132
FT VARIANT 145 145
FT VARIANT 153 153
FT VARIANT 203 203
FT VARIANT 261 261
FT VARIANT 302 302
FT VARIANT 316 316
FT VARIANT 337 337
FT VARIANT 367 367
FT VARIANT 422 422
FT VARIANT 429 429
FT VARIANT 433 433
SQ SEQUENCE 506 AA; 55093 MM; 687EJ310677FCF9CB CRC64;
Query Match 7.8%; Score 188.5; DB 1; Length 506;

```

```

Best Local Similarity 22.4%; Pred. No. 1.6e-05;
Matches 71; Conservative 52; Mismatches 139; Indels 55; Gaps 11;

OY 158 PRGNIGGK-----TLSVQLELODSTGWTCTVQNGKK--VERKIDVPCPAPBPK 208
DB 88 PRVTNSDATKRNNDMSIRISNTTPADAGVYCVKFKRKGDMERK-----SGPETH 141
OY 209 SCDKTHCPPELLGGPSVFLPPLPKKOTLMISRPEVTCVVAVDSH--EDPEVKFNWYVDG 266
DB 142 LTVSAKSPVLSPTV-----RAPEQVNTCTSHGSPNNISLKWIKNG 188
OY 267 VEVNNAKTPREQYNTSYVSVLYVHODWLNGKSKCKVSKALPA--PIEKTISKA 324
DB 189 NELSAQSVDPEEDNNYSINSTTKVLLATGDVHSGVCEVAHVTLQGGPPL----- 241
OY 325 KGGREQVYTLPSRREL-----KNQVSLTGLVKFSPSDIAVEESNGPENNKKTT 379
DB 242 RGTANLSETRVPTL-EITGSPAGNQVNTQVKNFPRHLQTLWLEGNMSRTAPAS 300
OY 380 PVLVDSDGSPFLYSKLTVDKSRMQGNVFCSVNHEALHNHYTQKSLSPGLQDDETC 439
DB 301 VFVFNKQGTNQNQSWFLVNSSAHREAVVLTQVCHDG-----QPAVSKNHTLEVSAPK 354
OY 440 EAQDEL-----DGLWTT 452
DB 355 DQDTGQTPGPNDSNMTS 371

RESULT 60
SHS1_RAT STANDARD; PRT; 509 AA.
AC P97710; O08951; O70426; Q9QW15;
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor
DE (SHP substrate-1) (SHPs-1) (inhibitory receptor SHP-1) (Signal-
DE regulatory protein alpha-1) (SIRP-alpha-1) (Brain Ig-1-like molecule
DE with tyrosine-based activation motifs) (Blt) (Macrophage fusion
DE receptor) (Macrophage membrane protein MFP150).
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 42-60; 68-91; 128-137; 150-158;
RP 174-189; 192-202; 204-212; 218-237; 259-270; 279-282; 405-415 AND
RP 446-453, N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, AND
RP INTERACTIONS WITH PTNP6 AND PTPN11.
RC TISSUE=Fetal fibroblast;
RX MEDLINE=97098667; PubMed=8943344;
RA Fujioke Y., Matczaki T., Noguchi T., Yamatsu A., Yamao T.,
RA Takahashi N., Tsuda M., Takada T., Kasuga M.,
RT "A novel membrane glycoprotein, SHP-1, that binds the SH2-domain-
RT containing protein tyrosine phosphatase SHP-2 in response to mitogens
RT and cell adhesion";
RL Mol. Cell. Biol. 16:6887-6899 (1996).
RN [2]
RP SEQUENCE FROM N.A., SEQUENCE OF 32-48 AND 446-453, FUNCTION, AND
RP PHOSPHORYLATION ON TYROSINE RESIDUES.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=97415431; PubMed=9271230;
RA Sano S.-I., Ohnishi H., Omori A., Hasegawa J., Kubota M.,
RT "Blt, an immune antigen receptor-like molecule in the brain.";
RL FEBS Lett. 411:327-334 (1997).
RN [3]
RP SEQUENCE FROM N.A., SEQUENCE OF 99-107; 128-149; 192-217; 405-417;
RP 419-429; 446-467 AND 496-506, N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RC STRAIN=Fischer 344; TISSUE=Macrophage;
RX MEDLINE=98449911; PubMed=9774638;
RA Saginario C., Sterling H., Beckers C., Kobayashi R., Solimena M.,
RA Ullu E., Vignery A.;

```



```

FT CONFLICT 450 450 R -> E (IN REF. 3; AA SEQUENCE).
FT CONFLICT 499 499 MISSING (IN REF. 3; AA SEQUENCE).
SQ SEQUENCE 509 AA: 55690 MW: 58BLFEOA4DD429F4 CRC64;

Query Match
Beet Local Similarity 7.8%; Score 188.5; DB 1; Length 509;
Matches 95; Conservative 65; Mismatches 148; Indels 109; Gaps 18;

QY 23 TQGNVVLGKGGDTLCTCTAQSQKSI-----QFHKNKNQIKILNQGSPFLTKGPKSLND 78
DB 37 TQADSVVAADSDATLCTCTVSSLTTPGPIKMFKEGGKRSPTYSFFIGEHPRTITVSD 96
QY 79 RADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLLVGLTANSPDTHLQGS 138
DB 97 -ATKNNM-----DFEICISNTPEDAGTYTC-VKQKGIVE-----PDTEIKSGG 141
QY 139 LTLTLESPPGSPSPSQCSPPKGNITQGGKITLSVSOLELQDSGTWTCTVLQNKQKVEFKID 198
DB 142 TTLVYLAKP-SSEPVSGPDSRG---SPQITVN-----FTC----- 172
QY 199 IVPCEAPPEKSCDKHTHCEPLLGGPSVFLFPKPKDITMISTRPVTCVWVDVSHEDPEV 258
DB 173 -----KS-----YGFSPR-----NI 182
QY 259 KFNWYDGEVHNATKPREEQNSTYEVSVLTV-LHQDWLNGEKYKCVSNKALPA-P 316
DB 183 TLKMKLKGDELISHLETTSKKS-NVSYNISTVSVKLSPEDIHSAVI-CEVAHTVLEGRP 240
QY 317 IEKITSKAKGPREQVYTLTPPSRDELTKNOVSLTCLVKGYPEDIVAVESNGQPPNN 375
DB 241 LNGTNFNNIIRVSP---TLKITQPLTPASQVNLTCQVKYPPALDLNMLNGNLSRT 297
QY 376 YKTPPVLDSDGSPFLYSKLTVDKSRMOQGNVFSQVWHE---ALAHNYTKSLIS 429
DB 298 DKPEHFTNDRDGTNYTSLFLVNSAHKEDVFTQVEHDSQPLTENHTVAPAFHS 354

RESULT 61
LAC1 MOUSE STANDARD; PRT; 105 AA.
AC P01843;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig lambda-1 chain C region.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=83014953; PubMed=6812053;
RA Seising E., Miller J., Wilson R., Storb U.;
RT "Evolution of mouse immunoglobulin lambda genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN 2;
RP SEQUENCE FROM N.A. (MOPC 315).
RX MEDLINE=81148806; PubMed=6259534;
RA Botchwell A.L.M., Paekind M., Schwartz R.C., Sorenshain G.E.,
RA Getter M.L., Baltimore D.;
RT "Dual expression of lambda genes in the MOPC-315 plasmacytoma.";
RL Nature 290:65-67(1981).
RN 3;
RP SEQUENCE FROM N.A. (S43).
RX MEDLINE=82220143; PubMed=6283385;
RA Botchwell A.L.M., Paekind M., Reth M., Imanishi-Kari T., Rajewsky K.,
RA Baltimore D.;
RT "Somatic variants of murine immunoglobulin lambda light chains.";
RL Nature 298:380-382(1982).
RN 4;
RP SEQUENCE (MYELOMA PROTEIN MOPC 104E).
RX MEDLINE=71107854; PubMed=5276767;
RA Appella E.;

```

```

RT "Amino acid sequences of two mouse immunoglobulin lambda chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 68:590-594(1971).
CC -1- MISCELLANEOUS: The MOPC 315 cell line produces 2 light chains, 1
CC normal lambda-2 chain and 1 abnormal lambda-1 chain that is
CC missing a large part of the V region. The C region sequence (shown
CC here) appears completely normal.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
DR EMBL; J00582; AA51636.1; -.
DR EMBL; J00587; AAB59672.1; -.
DR PIR; A93922; LIMS.
DR PDB; 1JNH; 06-FEB-02.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003066; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON_TER 1
FT DOMAIN 1
FT 6 100 IG-LIKE.
FT DISULFID 27 86
FT 104 104 INTERCHAIN (WITH HEAVY CHAIN).
FT CONFLICT 19 20 E -> TE (IN REF. 4).
FT CONFLICT 56 56 O -> E (IN REF. 4).
FT CONFLICT 75 75 MISSING (IN REF. 4).
FT CONFLICT 81 82 HS -> SH (IN REF. 4).
FT CONFLICT 85 85 S -> SS (IN REF. 4).
FT CONFLICT 96 96 E -> Q (IN REF. 4).
SQ SEQUENCE 105 AA; 11575 MW; A89F2B09BCFCA018 CRC64;

Query Match
Beet Local Similarity 7.4%; Score 177.5; DB 1; Length 105;
Matches 37; Conservative 18; Mismatches 43; Indels 5; Gaps 3;

QY 327 QPR-EPQVYTLTPPSRDELTKNOVSLTCLVKGYPEDIVAVESNGQPENNYKTPPYLD 384
DB 1 QPKSPSVTLTPPSSELETKNATLVCTITDFPGVTVDMKVDGTPVQGMETTPGSKQ 60
QY 385 SDGSPFLYSKLTVDKSRMOQGNVFSQVWHEALAHNYTKSLIS 427
DB 61 SNKYMASVYLLTLPARAMERHSSVCQVTHE---GHTVEKSLIS 100

RESULT 62
VCAL RAT STANDARD; PRT; 739 AA.
AC P29534;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=lung;
RX MEDLINE=92181437; PubMed=1371918;
RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wyck M.,
RA Buckley L., Miyake K., Kincaid P., Lobb R.;
RT "Cloning of murine and rat vascular cell adhesion molecule-1.";

```

```

RL Biochem. Biophys. Res. Commun. 183:163-169(1992).
CC - FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUCOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VIA4 ON LEUCOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUCOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
CC well as on macrophage-like and dendritic cell types in both normal
CC and inflamed tissue.
CC - SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-eb.ch).
CC -----
DR EMBL; M84488; AAA42332.1; -.
DR PIR; J50675; J50675.
DR HSSP; P19320; 1VCA.
DR InterPro; IPR003987; ICAM VCAM-1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig_5.
DR PRINTS; PR01472; ICAMVCAM1.
DR PRINTS; PR01474; VCAM1.
DR SMART; SM00408; IGC2_3.
DR PROSITE; PS50835; IG_LIKE; 5.
DR KMW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
DR Repeat; Signal.
KW SIGNAL.
FT 1 24 PROBABLE.
FT CHAIN 1 25 739 VASCULAR CELL ADHESION PROTEIN 1.
FT DOMAIN 25 698 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 699 720 POTENTIAL.
FT DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 119 25 IG-LIKE C2-TYPE 1.
FT DOMAIN 125 212 IG-LIKE C2-TYPE 2.
FT DOMAIN 223 309 IG-LIKE C2-TYPE 3.
FT DOMAIN 312 397 IG-LIKE C2-TYPE 4.
FT DOMAIN 408 506 IG-LIKE C2-TYPE 5.
FT DOMAIN 514 595 IG-LIKE C2-TYPE 6.
FT DOMAIN 601 682 IG-LIKE C2-TYPE 7.
FT DISULFID 47 95 BY SIMILARITY.
FT DISULFID 52 99 BY SIMILARITY.
FT DISULFID 137 195 BY SIMILARITY.
FT CARBOHYD 273 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 739 AA; 81246 MW; 5C608E5A1A1B100C CRC64;

```

```

Query Match 7.3%; Score 176.5; DB 1; Length 739;
Best Local Similarity 21.4%; Pred. No. 0.00016;
Matches 101; Conservative 71; Mismatches 169; Indels 131; Gaps 23;

```

```

QY 33 KQDVELTCTAOKSISQIHWKNSNOIKILNGSGFLVTGPKSLNDRADSRSLMDQGNF 92
DB 238 EGAAYMTCASBGLPAEPIFWSKK-----LDNGVLQL-----SGNA 274
QY 93 PLIKKLTEDSDTYCE-----VEDQKEVOLLV-----FGLTNSDTHLQGGSLTLT 142
DB 275 TLTLLMRHEDSGIYCEGVNLVGRDKTEVELVQGEKPTTVDISPSGVAAQVGSVLT 334
QY 143 LESPSSPSVQCRSPRGKNIQ-----GKLTYSOLELDPGCTWTCTVLONOKVVF 195
DB 335 CAAVGDSDFSFSWRTGTDSPLNGEVNDEGATSTLTLSPVGVDEHSHYLTCTVTCGRKLEK 394

```

```

QY 196 KIDIVCPAPRPSCDKTHTCPETLGGPSVFLPPPKKDTLMSRTEPVTCVVVDY--- 251
DB 395 TIQVEVYSPED-----PEIISGP-----LVHRPVTWCTVNVNVPF 433
QY 252 SHEDEP-----KENWYDVEVHNAKTKPREEOY-----NSTYVSVLTVLHODW 298
DB 434 DHLTEILKGETTLTNK--LREELGTSLSEKLSLEMFIPAEPTGVALVLAHSSQ 491
QY 299 LNKGEYKCKVSKNALPAIEKTISSAKQOPREPOVYTLR-PSRDLTKNQVSLTLVNGF 357
DB 492 ME-SEPKROSGTQTLVNVV-----APKEPTTVSPSPVPEE--GSPVNLTCSSDGF 539
QY 358 YPSDIAVME---SNGPENNYKTPVILDSGSEFLXSLTVDRSRWQGNVFCSVWH 414
DB 540 PPKKL--WSROLKNGELQ-----PLSQNTTSLFMFKM-----EDSGIYCEGIN 583
QY 415 EALHNHYQKSLT-----SPGLQL-----DETCAEADQDELGLM 450
DB 584 EA---GIKKSVELLIGSSKDQLTLVFPSPKSVKSGDVIYIISCTGSPVEIW 632

```

RESULT 63

ID	SRB2 HUMAN	STANDARD	PRT	387 AA.
AC	O9P1W8; O9P1W8; O9NOK8;			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Signal-regulatory protein beta-2 precursor (SIRP-beta-2) (SIRP-B2).			
GN	SIRPB2.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid=9606;			
RI	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RC	TISSUE=Placenta;			
RX	MEDLINE=21036165; PubMed=1185750;			
RA	Ichigotani Y., Matsuda S., Machida K., Oshina K., Iwamoto T.,			
RT	Yamaki K., Hayakawa T., Hamaguchi M.;			
RL	"Molecular cloning of a novel human gene (SIRP-B2) which encodes a new			
	member of the SIRP/SHPs-1 protein family.";			
	J. Hum. Genet. 45:378-382(2000).			
RP	[2]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=21638749; PubMed=11780052;			
RA	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggaley C.L.,			
RA	Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,			
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Frazer A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,			
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,			
RA	Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,			
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,			
RA	Lehvasaaho M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,			
RA	Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurtry A.A.,			
RA	Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,			
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,			
RA	Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,			
RA	Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,			
RA	Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,			
RA	Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,			
RA	Tracey A., Tsimanis A.C., Vaudin W., Wall M., Wallis J.M.,			
RA	Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,			
RA	Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,			
RT	Rogers J.;			
RL	"The DNA sequence and comparative analysis of human chromosome 20.";			
	Nature 414:865-871(2001).			

```

[3]
RN SEQUENCE FROM N.A. (ISOFORM 3).
RP TISSUE=Lung;
RC MEDLINE=22388257; PubMed=12477932;
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg H., Buetow K.H., Scheffer C.F., Bat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stacioner M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci P., Prange C.,
RA Rana S.S., Loguella N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- FUNCTION: Probable immunoglobulin-like cell surface receptor.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9P1W8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9P1W8-2; Sequence=VSP_007027;
CC Note=No experimental confirmation available;
CC Name=3;
CC IsoId=Q9P1W8-3; Sequence=VSP_007028;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Detected in liver, and at very low levels in
CC brain, heart, lung, pancreas, kidney, placenta and skeletal
CC muscle.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AB042624; BAA95692.1; -
CC EMBL, AL138804; CAC00474.1; -
CC EMBL, BC020629; AAH20629.1; ALT_INIT.
CC Genew, HGNC:15757; STRP2.
CC MIM, 605466; -
CC GO, GO:0007267; P:cell-cell signaling; TAS.
CC GO, GO:0007242; P:intracellular signaling cascade; TAS.
CC GO, GO:0008285; P:negative regulation of cell proliferation; TAS.
CC InterPro, IPR007110; Ig-like.
CC InterPro, IPR003597; Ig_C1.
CC InterPro, IPR003006; Ig_MHC.
CC InterPro, IPR003596; Ig_V.
CC Pfam, PF00047; Ig_3.
CC SMART, SM00407; IGc1. 2.
CC SMART, SM00406; IGv_1.
CC PROSITE, PS50835; IG_LIKE; 3.
CC PROSITE, PS00290; IG_MHC; FALSE_NEG.
CC Repeat, Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;
CC Alternative splicing.
CC SIGNAL 1 28 POTENTIAL.
CC CHAIN 29 387 SIGNAL-REGULATORY PROTEIN BETA-2.
CC DOMAIN 29 360 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 361 383 POTENTIAL.

```

```

FT DOMAIN 384 387 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 29 137 IG-LIKE V-TYPE.
FT DOMAIN 146 245 IG-LIKE C1-TYPE 1.
FT DOMAIN 252 340 IG-LIKE C1-TYPE 2.
FT DISULFID 53 119 POTENTIAL.
FT DISULFID 168 226 POTENTIAL.
FT DISULFID 271 329 POTENTIAL.
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 33 Missing (in isoform 2).
FT VARSPLIC 144 360 /FtId=VSP_007027.
FT VARSPLIC 263 263 Missing (in isoform 3).
FT CONFLICT 263 263 A -> V (IN REF. 2).
FT CONFLICT 286 286 L -> S (IN REF. 2).
FT CONFLICT 286 286 L -> S (IN REF. 2).
SQ SEQUENCE 387 AA; 42495 MW; F7F20C9F86E0E64B CRC64;

Query Match 7.2%; Score 175; DB 1; Length 387;
Best Local Similarity 22.0%; Pred. No. 9.2e-05;
Matches 100; Conservative 50; Mismatches 164; Indels 140; Gaps 17;

OY 6 PRRHLIVLQALIPAA-----TQGNKVLGKGGTVBELTTASQKKSIIQFHWKNSNQ 58
DB 13 PF--LLTLTLGLTEVAGEEELQMIQPEKLLVTVGKTATLHCTVT----- 56
OY 59 IKILNGSFLITGPSKLNRRADSRSLMDQGNP---LIINKLIEDSDTYICEVEDOK 115
DB 57 -----SLTPVGP-----VLMFVGVPGRRELY-----NOK 81
OY 116 EEVQLVFGTANSDDLHLOGSLTLTLFESPSPGSSPVSQCRSPRGKNIQSGKTLVSQLE 175
DB 82 E-----GHFPRVTVSDLTAKNNNDPFSIRISIT 110
OY 176 LQDSGTWTCVLQ--NOKYVEFKIDIVPCAPAPBPKCDKHTHCELGSPV-FLPPKP 232
DB 111 PADGTYCYVKFKRQSPENVEFK-----SGGTETWALAKP 146
OY 233 KDTLMIS-----RPPEVTCVAVDVSH--EDPEVKFNWVDGVEVHNATKTRBEDNYNSTR 286
DB 147 SAPVVLPAARATTPENTVSTCSHGSPRDIITKMKKNELSDPFTNVDPGQSAVYS 206
OY 287 VVSVLTVLHDWLNKGYKCVKSNKALPA-PIEKT--ISRAKQPRPOVYTLPPSRDEL 343
DB 207 IRSRARVLDPMDVRSGVIEVAHVTLQSGDPLAGTALSARINPPLLEV-TQGPMP--- 262
OY 344 TKNOVSLTCLVKGFPSPDIADVEMESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRWQ 403
DB 263 AGNQVNTVCQVRKFPQSLQTLWLENGNVCCRETASTLTENKQGTVMYMTSWFLVNIISDOR 322
OY 404 QGNVFGSCVWHEALHNHYTKSLISPGLOLDET 437
DB 323 DDVVLTCQVRHGD-----QLAVSRKLALEV 348

RESULT 64
LAC2_RAT STANDARD; PRT; 104 AA.
AC P20767;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-2 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87305594; PubMed=3114047;
RA Steen W.L., Hellman L., Petersson U.;
RT "The immunoglobulin lambda locus in rat consists of two C lambda

```

```

RT  genes and a single V lambda gene."
RL  Gene 55:75-84(1987).
CC  -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M22521; AAA1420.1; ALT_INIT.
DR  HSP: P01842; 2MCG.
DR  InterPro: IPR007110; IG-like.
DR  InterPro: IPR003597; IG_c1.
DR  InterPro: IPR003006; IG_MHC.
DR  Pfam: PF00047; Ig; 1.
DR  SMART: SM00407; IGc1; 1.
DR  PROSITE: PS50835; IG_LIKE; 1.
DR  PROSITE: PS00290; IG_MHC; 1.
DR  Immunoglobulin domain; Immunoglobulin C region.
KM  NON_TER 1 1
FT  DOMAIN 6 99
FT  DISULFID 27 85
FT  DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ  SEQUENCE 104 AA, 11318 MW, F087906DE43F7276 CRC64;

Query Match 7.1%; Score 172; DB 1; Length 104;
Best Local Similarity 35.3%; Pred. No. 2.8e-05;
Matches 36; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

OY 327 QPRE-POVYTLPPSDDELTKNOVSLCLVKGFPYSDIAVENSQNPENNYTTTPVLDS 385
DB 1 QPKSPFTLVFPSPSTBELQGNKATLVCLISDPYPSDVEYAMKNAPIISQGVDTANPTKQ 60
OY 386 DGSFPLYSKLTVDKSRMOQGNFSCSVMEALHNYTKSLS 427
DB 61 GNKYASSFLRLTAEQMRSRNSFTCVTHE--GNIVKSLS 99

RESULT 65
ID PGEM_HUMAN STANDARD; PRT; 4391 AA.
AC P98160; Q16287; Q9H3V5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112994; PubMed=1730768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecule, and epidermal growth factor."
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tian R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domain homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor."

```

```

RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJ51 TYR-1532.
RX MEDLINE=20553141; PubMed=1101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA Beighton P., Ben-Hamda C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Schwartz-Jampel syndrome (chondrodysdystrophic myotonia).";
RL Nat. Genet. 26:480-483(2000).
RN [4]
RP SEQUENCE OF 1016-1470 FROM N.A.
RC TISSUE=Colon.
RX MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1."
RL Genomics 10:673-680(1991).
RN [5]
RP SEQUENCE OF 890-1396 FROM N.A.
RC TISSUE=Fibroblastoma;
RX MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1-3p35 and identification of
RT a BamHI restriction fragment length polymorphism."
RL Genomics 11:389-396(1991).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessels S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter."
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT "Identification and quantification of N-linked glycoproteins using
RT hydrazone chemistry: versatile isotope labeling and mass spectrometry."
RL Nat. Biotechnol. 21:660-666(2003).
RN [8]
RP FUNCTION: This protein is an integral component of basement
RN membranes. It is responsible for the fixed negative electrostatic
RN charge and is involved in the charge-selective ultrafiltration
RN properties. It serves as an attachment substrate for cells.
RN [9]
RP SUBUNIT: Purified perlecan has a strong tendency to aggregate in
RN dimers or stellate structures. It interacts with other basement
RN membrane components such as laminin, prolargin and collagen type
RN IV.
RN [10]
RP SUBCELLULAR LOCATION: Extracellular.
RN [11]
RP TISSUE SPECIFICITY: Found in the basement membranes.
RN [12]
RP PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
RN AND O-LINKED OLIGOSACCHARIDES.
RN [13]
RP DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
RN syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
RN characterized by permanent myotonia (prolonged failure of muscle
RN relaxation) and skeletal dysplasia, resulting in reduced stature,
RN kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
RN [14]
RP SIMILARITY: Contains 4 LDL-receptor class A domains.
RN [15]
RP SIMILARITY: Contains 11 laminin EGF-like domains.
RN [16]
RP SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
RN [17]
RP SIMILARITY: Contains 3 laminin G-like domains.
RN [18]
RP SIMILARITY: Contains 4 EGF-like domains.
RN [19]
RP SIMILARITY: Contains 1 SEA domain.
RN [20]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X62515; CAA44373.1; -
 DR EMBL; M85289; AAS2700.1; -
 DR EMBL; AL445795; CAC18534.1; -
 DR EMBL; M64283; AAS2659.1; -
 DR EMBL; S76436; AAB21121.2; -
 DR EMBL; L22078; -; NOT_ANNOTATED_CDS.
 DR PIR; A38096; A38096.
 DR HSSP; P00740; IEDM.
 DR Sienna-2DPAGE; P98160; -
 DR Genew; HGNC:5273; HSPG2.
 DR MIM; 142461; -
 DR MIM; 255800; -
 DR InterPro; IPR008985; ConA_like_jec_g1.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003599; IG.
 DR InterPro; IPR003598; IG_c2.
 DR InterPro; IPR003596; IG_v.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00047; Ig; 22.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 7.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; IGL_recept; 4.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin_B; 3.
 DR SMART; SM00181; EGF; 15.
 DR SMART; SM00180; EGF_Lam; 12.
 DR SMART; SM00409; IG; 22.
 DR SMART; SM00408; IGC2; 21.
 DR SMART; SM00406; IGV; 7.
 DR SMART; SM00281; LamB; 3.
 DR SMART; SM00282; LamB; 3.
 DR SMART; SM00197; LDLa; 4.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 9.
 DR PROSITE; PS01186; EGF_2; 6.
 DR PROSITE; PS00026; EGF_3; 4.
 DR PROSITE; PS00835; IG_Like; 22.
 DR PROSITE; PS00025; IAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PROSITE; PS00024; SEA; 1.
 KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 4391
 FT DOMAIN 80 194
 FT DOMAIN 198 235
 FT DOMAIN 284 320
 FT DOMAIN 324 360
 FT DOMAIN 367 404
 FT DOMAIN 405 504
 FT DOMAIN 521 530
 FT DOMAIN 531 730
 FT DOMAIN 731 763
 FT DOMAIN 764 813

BASEMENT MEMBRANE-SPECIFIC HEPARAN
 SULFATE PROTEOGLYCAN CORE PROTEIN.
 SEA.
 LDL-RECEPTOR CLASS A 1.
 LDL-RECEPTOR CLASS A 2.
 LDL-RECEPTOR CLASS A 3.
 LDL-RECEPTOR CLASS A 4.
 IG-LIKE C2-TYPE 1.
 IG-LIKE C2-TYPE 1. (N-TERMINAL).
 LAMININ EGF-LIKE 1 (DOMAIN III A).
 LAMININ EGF-LIKE 1 (C-TERMINAL).
 LAMININ EGF-LIKE 2.

FT DOMAIN 814 871 LAMININ EGF-LIKE 3.
 FT DOMAIN 879 923 LAMININ EGF-LIKE 4 (INCOMPLETE).
 FT DOMAIN 924 933 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 934 1125 LAMININ DOMAIN IV 2 (DOMAIN III B).
 FT DOMAIN 1126 1158 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 1159 1208 LAMININ EGF-LIKE 6.
 FT DOMAIN 1209 1265 LAMININ EGF-LIKE 7.
 FT DOMAIN 1275 1324 LAMININ EGF-LIKE 8.
 FT DOMAIN 1325 1334 LAMININ EGF-LIKE 9 (N-TERMINAL).
 FT DOMAIN 1335 1529 LAMININ EGF-LIKE 9 (DOMAIN III C).
 FT DOMAIN 1530 1562 LAMININ EGF-LIKE 9 (C-TERMINAL).
 FT DOMAIN 1563 1612 LAMININ EGF-LIKE 10.
 FT DOMAIN 1613 1670 LAMININ EGF-LIKE 11.
 FT DOMAIN 1677 1771 IG-LIKE C2-TYPE 2.
 FT DOMAIN 1772 1865 IG-LIKE C2-TYPE 3.
 FT DOMAIN 1866 1955 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1956 2051 IG-LIKE C2-TYPE 5.
 FT DOMAIN 2052 2151 IG-LIKE C2-TYPE 6.
 FT DOMAIN 2152 2244 IG-LIKE C2-TYPE 7.
 FT DOMAIN 2245 2340 IG-LIKE C2-TYPE 8.
 FT DOMAIN 2341 2436 IG-LIKE C2-TYPE 9.
 FT DOMAIN 2437 2533 IG-LIKE C2-TYPE 10.
 FT DOMAIN 2534 2629 IG-LIKE C2-TYPE 11.
 FT DOMAIN 2630 2726 IG-LIKE C2-TYPE 12.

Query Match 7.0%; Score 168.5; DB 1; Length 4391;
 Best Local Similarity 20.3%; Pred. No. 0.0053;
 Matches 96; Conservative 66; Mismatches 192; Indels 119; Gaps 21;

Qy 33 KGDVETLTCTASQKSKIQFHWKNSNQIKLGNQSGFTTKPSPKLNDRADRRSLMDQGNF 92
 Db 2740 EGETLIDLCVPGQAHQVTH-----KKGSL-PSHQTRGSRLLR----- 2779
 Qy 93 PLIKNKIEDSPYICEVEDQK---EEVQLVFGLTANSDTL----- 132
 Db 2780 ---LHVSPPADSGERYCRVWGSSGPLEASVLTIEASGSAVHVAPGAPPIRIEPPSS 2836
 Qy 133 -LLQGSLTTLSPSSPSVQCRSPRGKNIQG-----GKTLVSQLELDQSGTWTCV 186
 Db 2837 RVAGGQTLDKC-VVPGQAHQVTHMKRGNLPAHQVHPLRLRNQVSPADSEYSCQV 2895
 Qy 187 LQNKQVEFKIDV----PCAPPEKSCDKHTTCPELLGSPVFLFPPEKDTLMISRT 241
 Db 2896 TGSSGTLASGLVLTIEPSSGPPIAP-----GLAQPIYIEASSSHVTEGQT 2941
 Qy 242 PEVTCVVVDVSHEDPEVKFMVVDGVEVHNAKTRPREQVNSTYRVVSVLTVLHQDLNG 301
 Db 2942 LDLCVVPQGNH---AQVTWYRG---GSLPARHQHGSQURLHLVSPA-----DS 2986
 Qy 302 KEVKCVSNKALP---APIEKTSKAKGP---REPQVTLPPSRDELTKNOVSLTCLV- 354
 Db 2987 GEVYCRASGPGEQKASFTVTVPSESSRYRLRSPVTSIDPSSSTVQGGDASFKCLIH 3046
 Qy 355 KGRPPSDIAVEMESNGCP-ENNTKTPPVLDSDGSFFLYSKLTVYDKSRMOQGNVFC-- 410
 Db 3047 DGAAP--ISLEWTKRQLEDELDNHSIP-----NGSI-----ITVGRPPNHGTYRCVAS 3094
 Qy 411 -----SWHBLHNHYTKSLSLSP-----GLQDETCAEADQDGLDGLMT 451
 Db 3095 NAYGASVNVLSVHGPPITVSLPEGFVWVKVGAATLBCVASAGEPPRSARMT 3147

RESULT 66
 CD22_HUMAN STANDARD; PRT; 847 AA.
 AC P20273; O95699; O95701; O95702; O95703; Q01665; Q92872; Q92873;
 AC Q9UQA7; Q9UQA8; Q9UQA9; Q9UQA0; Q9Y2A6;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE B-cell receptor CD22 precursor (Ileu-14) (B-lymphocyte cell adhesion
 molecule) (Bt-Cam) (Siglec-2).
 GN CD22.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OK NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM CD22-BETA), AND VARIANT HIS-639.
 RC TISSUE=tonal;
 RX MEDLINE=91086838; PubMed=1985119;
 RA Wilson G.L., Fox C.H., Fauci A.S., Kehrl J.H.;
 RT "CD4 cloning of the B cell membrane protein CD22: a mediator of B-B
 cell interactions.";
 RL J. Exp. Med. 173:137-146(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).
 RX MEDLINE=93267103; PubMed=8496602;
 RA Wilson G.L., Najfeld V., Kozlow E., Menniger J., Ward D.,
 RA Kehrl J.H.;
 RT "Genomic structure and chromosomal mapping of the human CD22 gene.";
 RL J. Immunol. 150:5013-5024(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).
 RX Lamerdin J.E., McCreedy P., Adamson A.W., Burkhardt-Schulz K.,
 RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Gaines J.,
 RA Danganan L., Bruce R., Quan G., Montgomery M., Ow D.,
 RA Kobayashi A., Olsen A.O., Carrano A.V.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM CD22-BETA).
 RX MEDLINE=90231465; PubMed=1691828;
 RA Stamenkovic I., Seed B.;
 RT "The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion.";
 RL Nature 345:74-77(1990).
 RN [5]
 RP SEQUENCE OF 13-137; 139-239; 241-328 AND 418-502 FROM N.A., AND
 RP VARIANTS THR-24; GLU-152; LYS-203; GLY-664; CYS-669 AND ASP-745.
 RX MEDLINE=99180618; PubMed=10079291;
 RA Harita Y., Tsuchiya N., Matsushita M., Shiohara M., Hagiwara K.,
 RA Tokunaga K.;
 RT "Identification of the gene variations in human CD22.";
 RL Immunogenetics 49:280-286(1999).
 RN [6]
 RP SIALIC ACID BINDING.
 RX MEDLINE=93316636; PubMed=8463235;
 RA Powell L.D., Sgroi D., Sjoberg E.R., Stamenkovic I., Varki A.;
 RT "Natural ligands of the B cell adhesion molecule CD22 beta carry
 N-linked oligosaccharides with alpha-2,6-linked sialic acids that are
 required for recognition.";
 RL J. Biol. Chem. 268:7019-7027(1993).
 RN [7]
 RP INTERACTION WITH PTNP6.
 RX MEDLINE=95343349; PubMed=7618087;
 RA Doody G.M., Justement L.B., Delibrias C.C., Matthews R.J., Lin J.,
 RA Thomas M.L., Fearon D.T.;
 RT "A role in B cell activation for CD22 and the protein tyrosine
 phosphatase SHP.";
 RL Science 265:242-244(1995).
 RN [8]
 RP INTERACTION WITH LYN, SYK AND PIK3R1/PIK3R2.
 RX MEDLINE=96257803; PubMed=8647200;
 RA Tusciano J.M., Engel P., Tedder T.F., Agarwal A., Kehrl J.H.;
 RT "Involvement of p72syk kinase, p53/56lyn kinase and phosphatidylinositol-3 kinase in signal transduction via the human B lymphocyte
 antigen CD22.";
 RL Eur. J. Immunol. 26:1246-1252(1996).
 RN [9]
 RP INTERACTION WITH PTNP6, SYK AND PLCG1.
 RX MEDLINE=96195207; PubMed=8627166;
 RA Law C.L., Sidorenko S.P., Chandran K.A., Zhao Z., Shen S.H.,
 RA Fischer E.H., Clark E.A.;
 RT "CD22 associates with protein tyrosine phosphatase 1C, Syk, and
 phospholipase C-gamma(1) upon B cell activation.";
 RL J. Exp. Med. 183:547-560(1996).
 RN [10]

RP REVIEW.
 RX MEDLINE=97288746; PubMed=9143697;
 RA Tedder T.F., Tusciano J., Sato S., Kehrl J.H.;
 RT "CD22, a B lymphocyte-specific adhesion molecule that regulates
 antigen receptor signaling.";
 RL Annu. Rev. Immunol. 15:481-504(1997).
 CC -1- FUNCTION: Mediates B-cell B-cell interactions. May be involved in
 the localization of B-cells in lymphoid tissues. Binds sialylated
 glycoproteins, one of which is CD45. Preferentially binds to
 alpha2,6-linked sialic acid. The sialic acid recognition site can
 be masked by cis interactions with sialic acids on the same cell
 surface. Upon ligand induced tyrosine phosphorylation in the
 immune response seems to be involved in regulation of B cell
 antigen receptor signaling. Play a role in positive regulation
 through interaction with Src family tyrosine kinases and may also
 act as an inhibitory receptor by recruiting cytoplasmic
 phosphatases via their SH2 domains that block signal transduction
 through dephosphorylation of signaling molecules.
 CC -1- SUBUNIT: Predominantly monomer of isoform CD22-beta. Also found as
 heterodimer of isoform CD22-beta and a shorter isoform. Interacts
 with PTNP6/SHP-1, LYN, SYK, PIK3R1/PIK3R2 and PLCG1 upon
 phosphorylation. Interacts with GRB2, INPP5D and SHC1 upon
 phosphorylation (By similarity). May form a complex with
 INPP5D/SHP, GRB2 and SHC1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Comment=Additional isoforms seem to exist;
 Name=CD22-beta;
 Name=CD22-alpha;
 Name=CD22-1; Sequence=Displayed;
 Name=CD22-2; Sequence=VSP_002531;
 CC -1- TISSUE SPECIFICITY: B lymphocytes.
 CC -1- DOMAIN: Contains 4 copies of a cytoplasmic motif that is referred
 to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).
 CC This motif is involved in downmodulation of cellular responses.
 CC The phosphorylated ITIM motif binds to the SH2 domain of
 PTNP6/SHP-1.
 CC -1- PTM: Phosphorylated both on threonine/serine and tyrosine
 residues.
 CC -1- PTM: Phosphorylated on tyrosine residues by LYN (By similarity).
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
 (sialic acid binding Ig-like lectin) family.
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -1- CAUTION: Ref.4 sequence differs from that shown due to a
 frameshift in position 806.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD22 entry;
 WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd22.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X59350; CAA42006.1; -;
 CC EMBL, U62631; AAB06448.1; -;
 CC EMBL, U62631; AAB06449.1; -;
 CC EMBL, X52785; CAA36988.1; ALT_FRAME.
 CC EMBL, AB012996; BAA36565.1; -;
 CC EMBL, AB012997; BAA36566.1; -;
 CC EMBL, AB012998; BAA36567.1; -;
 CC EMBL, AB012999; BAA36568.1; -;
 CC EMBL, AB013000; BAA36569.1; -;
 CC EMBL, AB013002; BAA36571.1; -;
 CC EMBL, AB013003; BAA36572.1; -;
 CC EMBL, AB013004; BAA36573.1; -;
 CC EMBL, AB013006; BAA36575.1; -;
 CC GeneW; HGNC:1643; CD22.
 CC MIM; 107266; -;

```

DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; Ig; 6.
DR SMART; SM00408; IGC2; 4.
DR PROSITE; PS00835; IG_LIKE; 6.
KW Cell adhesion; Lectin; Antigen; Transmembrane; Signal; Glycoprotein;
KW Immunoglobulin domain; Repeat; Phosphorylation; Alternative splicing;
KW Polymorphism.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 847 B-CELL RECEPTOR CD22.
FT DOMAIN 20 687 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 688 706 POTENTIAL.
FT DOMAIN 707 847 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 138 IG-LIKE V-TYPE.
FT DOMAIN 143 235 IG-LIKE C2-TYPE 1.
FT DOMAIN 242 326 IG-LIKE C2-TYPE 2.
FT DOMAIN 331 416 IG-LIKE C2-TYPE 3.
FT DOMAIN 419 500 IG-LIKE C2-TYPE 4.
FT DOMAIN 505 582 IG-LIKE C2-TYPE 5.
FT DOMAIN 593 676 IG-LIKE C2-TYPE 6.
FT SITE 760 765 ITIM MOTIF 1.
FT SITE 794 799 ITIM MOTIF 2.
FT SITE 820 825 ITIM MOTIF 3.
FT SITE 840 845 ITIM MOTIF 4.
FT DISULFID 39 167 BY SIMILARITY.
FT DISULFID 44 102 BY SIMILARITY.
FT DISULFID 161 219 BY SIMILARITY.
FT DISULFID 265 309 BY SIMILARITY.
FT DISULFID 353 396 BY SIMILARITY.
FT DISULFID 442 484 BY SIMILARITY.
FT DISULFID 529 571 BY SIMILARITY.
FT DISULFID 616 659 BY SIMILARITY.
FT MOD_RES 762 762 PHOSPHORYLATION
(INVOLVED IN BINDING TO SYK)
FT MOD_RES 807 807 (BY SIMILARITY).
PHOSPHORYLATION
(INVOLVED IN BINDING TO GRB2)
FT MOD_RES 822 822 (BY SIMILARITY).
PHOSPHORYLATION
(INVOLVED IN BINDING TO SYK)
FT MOD_RES 842 842 (BY SIMILARITY).
PHOSPHORYLATION
(INVOLVED IN BINDING TO SYK, PLCG2 AND
PIK3R1/PIK3R2) (BY SIMILARITY).
FT CARBOHYD 67 67 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.9%; Score 167; DB 1; Length 847;
Best Local Similarity 21.2%; Pred. No. 0.00085;
Matches 95; Conservative 70; Mismatches 154; Indels 130; Gaps 23;
32 KKGDVVELCTASQKKSIOFHMKNSTQIKINGQSFITKGPSKLNDRASRLMDQGN 91
256 REGDSVMTCEVSS-----SNPEYTV-----SWLKDG-----TSLKQNT 291
92 PPLTIKNIKIDSDTYICEVED-----QKEEVQLIVFGLTANSDTHLL-----QGQS 138
-292 FTLMIREYTKQSGKYCCQVENDVQGRSEEVFLQVYAPRPSTVQILHSDPAVSGQVEF 351
139 LTLTLESPPSSSPQSCSPRGKNIG--GKTLISQLELDSDGWTCTVQONQKVEFK 196
352 LCMSTLANP--LPNTYTYTHNGKEMQGTKEKVHLPKILPMHAGTYSQ-VAEN----- 400
197 IDIVCPAPEBKSCDKHTCEPLG---GPSVFL---FPPKPKDTLMISRTP-----EVT 245
401 -----ILCTGORGCAELDVQYPRKXTTVYIGNMPPIREGDTV 438
246 CVVVDVSHEDPEV-KENNYVDGVEVHNAKTYPRERQYNSTYRVSVLTVLHQDLINGKEY 304
439 TLSGYNNSNPSVTRYEW-----KPHGAMBEPS-----LGVLTQIVGWMDN-TTI 482
QY KCKVSNK--ALPAPIEKTIKAKQPRRPQYVYTLPPSDELTKQVSLTCLVKGFYFEDI 362
DB ACAACNSWCSMASPVALLNVQYA---PRQVRVRKIKPIELSEIHSNGSVSLQCPSSSHPKEV 539

```

```

QY 363 AVENESNG-----QENNNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFCSEVHEAL 417
DB 540 QFFENKGRILGKESQNFPSISP--EDAGSY-----SCWVANSIGQTASKXKMTL 587
QY 418 HNNYTKSL--SUSPGQLDDE-----TC 438
DB 588 EVLYAPRRLRVSMSPGDQVMEGKSATLTC 616

RESULT 67
SRBI_HUMAN STANDARD; PRT; 398 AA.
AC 000241; O8TB12; Q9H1U5; Q9Y4V0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Signal-regulatory protein beta-1 precursor (SIRP-beta-1).
GN SIRPB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Placenta;
RX MEDLINE=97215901; Pubmed=9062191;
RA Khaitonenkov A., Chen Z., Sures I., Wang H., Schilling J.,
RT "A family of proteins that inhibit signalling through tyrosine kinase
RT receptors."
RL Nature 386:181-186(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; Pubmed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Cobby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhana P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharasleho M.H., Leverhna M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McClay K., McMurtry A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showmken R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Walli M., Wallis J.M.,
RA Whitehead S.L., Whitlaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20."
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Skin;
RX MEDLINE=22386257; Pubmed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,
RA Dlatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toehiyuki S., Carninci P., Prange C.,

```

RA Raha S.S., Loquejano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gnatrante P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.M.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.T., Skalska U., Smalhus D.E.,
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RN INTERACTION WITH TYROBP AND SYK.
 RP PubMed:1169422;
 RA Tomaseello E., Cant C., Buehring H.-J., Vely F., Andre P., Seifert M.,
 RA Ullrich A., Vivier E.;
 RL "Association of signal-regulatory proteins beta with KARAP/DAP-12.";
 RL Eur. J. Immunol. 30:2147-2156 (2000).
 RN [5]
 RN INTERACTION WITH TYROBP, N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
 RP MEDLINE:2007221; PubMed:10604985;
 RA Dietrich J., Cella M., Seifert M., Buehring H.-J., Colonna M.;
 RT "Signal-regulatory protein beta 1 is a DAP12-associated activating
 RT receptor expressed in myeloid cells.";
 RL J. Immunol. 164:9-12 (2000).
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor involved in
 CC the negative regulation of receptor tyrosine kinase-coupled
 CC signaling processes. Participates also in the recruitment of
 CC tyrosine kinase SYK.
 CC -1- SUBUNIT: Interacts with TYROBP. This interaction results in the
 CC recruitment of SYK.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O00241-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O00241-2; Sequence=VSP_007026;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Detected in monocytes and dendritic cells.
 CC -1- PTM: N-glycosylated.
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y10376; CAA71404.1; -;
 DR EMBL: AL049634; CAB46661.2; -;
 DR EMBL: AL138804; CAC17540.1; -;
 DR EMBL: BC025286; AAH25286.1; -;
 DR Genew; HGNC:15928; SIRPB1.
 DR MIM: 603889; -;
 DR GO: GO:0005887; C:integral to plasma membrane; TAS.
 DR GO: GO:0007166; P:cell surface receptor linked signal transdu. .; TAS.
 DR InterPro: IPR007110; IG_1-like.
 DR InterPro: IPR003597; IG_C1.
 DR Pfam: PF0047; Ig_3.
 DR SMART: SMO0407; IGc1. 2.
 DR PROSITE: PSS0835; IG_Like; 3.
 KW Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;
 KW Alternative splicing.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 398 SIGNAL-REGULATORY PROTEIN BETA-1.
 FT DOMAIN 27 371 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 372 392 POTENTIAL.

FT DOMAIN 393 398 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 27 136 IG-LIKE V-TYPE.
 FT DOMAIN 147 246 IG-LIKE C1-TYPE 1.
 FT DOMAIN 253 347 IG-LIKE C1-TYPE 2.
 FT DISULFID 54 120 POTENTIAL.
 FT DISULFID 169 227 POTENTIAL.
 FT DISULFID 272 320 POTENTIAL.
 FT CARBOHYD 244 244 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 269 269 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 291 291 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT VARSPLIC 145 361 Missing (in isoform 2).
 FT CONFLICT 53 53 R -> H (IN REF. 2; CAC17540).
 FT CONFLICT 102 102 D -> N (IN REF. 1).
 FT CONFLICT 229 229 M -> I (IN REF. 2; CAB46661).
 FT CONFLICT 363 363 P -> A (IN REF. 2).
 SQ SEQUENCE 398 AA; 43255 MW; A2AA08FEBB2BC52B CRC64;
 Query Match 6.8%; Score 165; DB 1; Length 398;
 Best Local Similarity 24.4%; Pred. No. 0.00045;
 Matches 76; Conservative 49; Mismatches 139; Indels 48; Gaps 13;
 QY 136 GQSLTLTLESPPGSSPVQCRSPRGKNIQGKTSVSQLQDSGTWTCTVLQ--NOKV 193
 DB 72 GAGRELIYNQKEGHPRTVTVSELTKRNNLDPFISISITPADAGTYCVAFKGSPPDV 131
 QY 134 EFKIDIVCPAPBPSCCKHTTCPELLGSGSVFLFPKPKOTLMSTRPEVTGVVDVSH 253
 DB 132 EFK-----SGAGTELSTVAKSPAP-VWGGPAV-----RATPEHTVSPFCSSH 172
 QY 254 --EDEPEFNNYVDQVEVHNAKTK--PRE-----QYNSTVAVSVLTVLQDMLNGKEYK 305
 DB 173 GSPSPDITLKKFKGNELSDPQTNVDPAGDSVSIHSTAVVLTGCVHSGV----- 226
 QY 306 CKVSNKALPA-PIEKT--ISRAKGPREDQVYTLPPSDELTQKQVSLTCLVKGFPSDI 362
 DB 227 CEMAHITLQGPPLRLGTALNSEAIRVPTLEV-TQOPMAE---NQANVTCQVSNFYPRGL 282
 QY 363 AVESNSQCPENNYTTPPVLDSDGSFPLYSKLYTDSKRMQGNVFGSCVWHEALHNHT 422
 DB 283 QLTWENENSVRTEPASTLIENKQGTVMWMLVNTCAHDDVLTQVHDG----- 336
 QY 423 QKSLSLSPGLQ 434
 DB 337 QQAVSKSYALSI 348
 Db
 RESULT 68
 UN99_CABE1 STANDARD; PRT; 6632 AA.
 AC 001761; Q17362;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
 GN UNC-89 OR C09P1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN=Bristol N2;
 RX MEDLINE=96180278; PubMed=8603916;
 RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
 RT assembly, encodes a giant modular protein composed of Ig and signal
 RT transduction domains".
 RL J. Cell Biol. 132:835-848 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T., Wilson R.;

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Materson R.;
CC Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Structural component of the muscle M-line. Myofibril
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 5 RCSD domains.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U33058; AAB00542.1; -.
CC EMBL: AF003131; AAB54132.2; -.
CC PDB: 1PHO; 20-DEC-00.
CC Wormpep; C09D1.1; CE30426.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003006; IG_LMC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhogGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; Ig; 47.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhogGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGc2; 23.
DR SMART; SM00325; RhogGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0010; DH_2; 1.
DR PROSITE; PSS0835; IG_LIKE; 49.
DR PROSITE; PSS0003; PH DOMAIN; 1.
DR PROSITE; PSS0002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1882 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.

FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6130 6230 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 IG-LIKE C2-TYPE 48.
FT DOMAIN 6413 6502 FIBRONECTIN TYPE-III.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DOMAIN 6568 6621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFICT 2137 2137 A -> P (IN REF. 1).
FT CONFICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFICT 2258 2258 A -> P (IN REF. 1).
FT CONFICT 2284 2284 E -> G (IN REF. 1).
FT CONFICT 2297 2297 M -> I (IN REF. 1).
FT CONFICT 3531 3531 A -> G (IN REF. 1).
FT CONFICT 3884 3888 DAGEY -> RRRI (IN REF. 1).
FT CONFICT 3929 3929 A -> V (IN REF. 1).
FT CONFICT 5134 5134 A -> P (IN REF. 1).
FT CONFICT 5145 5145 T -> S (IN REF. 1).
FT CONFICT 5185 5185 G -> A (IN REF. 1).
FT CONFICT 5199 5199 K -> N (IN REF. 1).
FT CONFICT 5202 5202 L -> F (IN REF. 1).
FT CONFICT 5213 5213 F -> L (IN REF. 1).
FT CONFICT 6178 6178 A -> G (IN REF. 1).
FT CONFICT 6268 6268 K -> E (IN REF. 1).
SO SEQUENCE 6632 AA; 73165 MW; 262D3EDD52960E89 CRC64;
Query Match 6.8%; Score 164; DB 1; Length 6632;

```

RESULT 69
ID LAC3_MOUSE STANDARD; PRT; 104 AA.
AC P01845;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE I9 lambda-3 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83014953; PubMed=6812053;
RA Seising E., Miller J., Wilson R., Storb U.;
RT "Evolution of mouse immunoglobulin lambda genes."
RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN [2]
RP SEQUENCE OF I-18 (MYELOMA PROTEIN CBPC-49).
RA Breyer R.M., Sauer R.T., Eisen H.N.;
RT "The variable region of mouse lambda-3 chains."
RL ICN UCLA Symp. Mol. Cell. Biol. 20:105-110(1981).
RN [3]
RP SEQUENCE OF I0-104 (MYELOMA PROTEIN CBPC-49 AND MONOCLONAL ANTIBODY
RP 8-47).
RX MEDLINE=81223782; PubMed=6165998;
RA Azuma T., Steiner L.A., Eisen H.N.;
RT "Identification of a third type of lambda light chain in mouse
RT immunoglobulins."
RL Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).
RN [4]
RP SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; J00585; AAB59670.1; --

DR PIR; B93922; L3MS.

DR HSSP; P01842; 2MCG.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003597; Ig_c1.

DR InterPro; IPR003006; Ig_MHC.

DR Pfam; PF00047; Ig; 1.

DR SMART; SM00407; Igcl; 1.

DR PROSITE; PS50835; IG_LIKE; 1.

DR PROSITE; PS00290; IG_MHC; 1.

DR Immunoglobulin domain; Immunoglobulin C region.

FT NON TER 1 1

FT DOMAIN 6 99 IG-LIKE.

FT DISULFD 27 85

FT DISULFD 103 103 INTERCHAIN (WITH HEAVY CHAIN).

SQ SEQUENCE 104 AA; 11371 MW; 83CECDAA348EFL CRC64;

```

RESULT 70
PCBM_MOUSE
ID PCBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPc) (Perlecan) (PLC).
DE HSPG2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=92078153; Pubmed=1744087;
RA Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule.";
RL J. Biol. Chem. 266:22939-22947 (1991).
RL [2]
RN SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=89034110; Pubmed=2972708;
RX Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 263:16379-16387 (1988).
RL -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC

```

```

CC      membrane components such as laminin, prolargin and collagen type
CC      IV.
CC      -1- SUBCELLULAR LOCATION: Extracellular.
CC      -1- TISSUE SPECIFICITY: Found in the basement membranes.
CC      -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC      AND O-LINKED OLIGOSACCHARIDES.
CC      -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC      -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC      -1- SIMILARITY: Contains 3 laminin IV domains.
CC      -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC      -1- SIMILARITY: Contains 3 laminin G-like domains.
CC      -1- SIMILARITY: Contains 1 EGF-like domain.
CC      -1- SIMILARITY: Contains 1 SEA domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M77174; AAA39911.1; -.
CC      EMBL; J04054; AAA39899.1; -.
CC      EMBL; J04055; AAA39912.1; -.
CC      PIR; S18252; S18252.
CC      PDB; 1GL4; 28-NOV-01.
CC      MGD; MGI:96257; Hap2.
CC      GO; GO:0005604; C:basement membrane; IDA.
CC      GO; GO:0008104; P:protein localization; IMP.
CC      DR      InterPro; IPR008985; Conn_like_jec_g1.
CC      DR      InterPro; IPR000742; EGF_2.
CC      DR      InterPro; IPR006209; EGF-like.
CC      DR      InterPro; IPR007110; IG-like.
CC      DR      InterPro; IPR003598; IG_c2.
CC      DR      InterPro; IPR000034; Laminin_B.
CC      DR      InterPro; IPR002049; Laminin_EGF.
CC      DR      InterPro; IPR001791; Laminin_G.
CC      DR      InterPro; IPR002172; LDL_receptor_A.
CC      DR      InterPro; IPR000082; SEA_domain.
CC      DR      Pfam; PF00008; EGF; 4.
CC      DR      Pfam; PF00047; IG; 15.
CC      DR      Pfam; PF00052; laminin_B; 3.
CC      DR      Pfam; PF00053; laminin_EGF; 7.
CC      DR      Pfam; PF00054; laminin_G; 3.
CC      DR      Pfam; PF00057; ldl_recept_a; 4.
CC      DR      Pfam; PF01390; SEA; 1.
CC      DR      PRINTS; PR00261; LDLRECEPTOR.
CC      DR      PRODOM; PD003031; Laminin_B; 3.
CC      DR      SMART; SM00180; EGF_Lam; 7.
CC      DR      SMART; SM00408; IGC2; 14.
CC      DR      SMART; SM00281; LamB; 3.
CC      DR      SMART; SM00282; LamG; 3.
CC      DR      SMART; SM00192; LDLA; 4.
CC      DR      SMART; SM00200; SEA; 1.
CC      DR      PROSITE; PS00022; EGF_1; 8.
CC      DR      PROSITE; PS01186; EGF_2; 5.
CC      DR      PROSITE; PS50026; EGF_3; 4.
CC      DR      PROSITE; PS50835; IG_LIKE; 15.
CC      DR      PROSITE; PS50025; LAM_G_DOMAIN; 3.
CC      DR      PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
CC      DR      PROSITE; PS01209; LDLRA_1; 4.
CC      DR      PROSITE; PS50068; LDLRA_2; 4.
CC      DR      PROSITE; PS50024; SEA; 1.
CC      KW      Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
CC      Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
CC      Extracellular matrix; EGF-like domain; 3d-structure.
CC      SIGNAL      1      21      POTENTIAL.
CC      CHAIN      22      3707      BASEMENT MEMBRANE-SPECIFIC HEPARAN
CC      FT      DOMAIN      80      194      SEA.
CC      FT      DOMAIN      195      234      LDL-RECEPTOR CLASS A 1.
CC      FT      DOMAIN      281      319      LDL-RECEPTOR CLASS A 2.

```

```

FT      320      359      LDL-RECEPTOR CLASS A 3.
FT      360      403      LDL-RECEPTOR CLASS A 4.
FT      404      504      IG-LIKE C2-TYPE 1.
FT      504      530      LAMININ EGF-LIKE 1. (N-TERMINAL).
FT      531      730      LAMININ EGF-LIKE 1. (DOMAIN III A).
FT      731      763      LAMININ EGF-LIKE 1. (C-TERMINAL).
FT      764      813      LAMININ EGF-LIKE 2.
FT      814      871      LAMININ EGF-LIKE 3.
FT      872      923      LAMININ EGF-LIKE 4. (INCOMPLETE).
FT      924      933      LAMININ EGF-LIKE 5. (N-TERMINAL).
FT      934      1125      LAMININ DOMAIN IV 2. (DOMAIN III B).
FT      1126      1158      LAMININ EGF-LIKE 5. (C-TERMINAL).
FT      1159      1208      LAMININ EGF-LIKE 6.
FT      1209      1265      LAMININ EGF-LIKE 7.
FT      1275      1324      LAMININ EGF-LIKE 8.
FT      1325      1334      LAMININ EGF-LIKE 9. (N-TERMINAL).
FT      1335      1529      LAMININ EGF-LIKE 9. (N-TERMINAL).
FT      1530      1562      LAMININ EGF-LIKE 9. (DOMAIN III C).
FT      1563      1612      LAMININ EGF-LIKE 10.
FT      1613      1670      LAMININ EGF-LIKE 11.
FT      1677      1771      IG-LIKE C2-TYPE 2.
FT      1772      1865      IG-LIKE C2-TYPE 2.
FT      1866      1954      IG-LIKE C2-TYPE 4.
FT      1955      2049      IG-LIKE C2-TYPE 5.
FT      2050      2148      IG-LIKE C2-TYPE 6.
FT      2149      2244      IG-LIKE C2-TYPE 7.
FT      2245      2343      IG-LIKE C2-TYPE 8.
FT      2344      2436      IG-LIKE C2-TYPE 9.
FT      2437      2532      IG-LIKE C2-TYPE 10.
FT      2533      2619      IG-LIKE C2-TYPE 11.
FT      2620      2720      IG-LIKE C2-TYPE 12.
FT      2721      2809      IG-LIKE C2-TYPE 13.
FT      2810      2895      IG-LIKE C2-TYPE 14.
FT      2896      2980      IG-LIKE C2-TYPE 15.
FT      2984      3162      LAMININ G-LIKE 1.
FT      3163      3241      LAMININ G-LIKE 2.
FT      3242      3425      LAMININ G-LIKE 3.
FT      3426      3518      HEPARAN SULFATE (POTENTIAL).
FT      3519      3705      HEPARAN SULFATE (POTENTIAL).
FT      3706      67      HEPARAN SULFATE (POTENTIAL).
FT      68      71      HEPARAN SULFATE (POTENTIAL).
FT      72      73      MEDIATES MOTOR NEURON ATTACHMENT
FT      74      76      (POTENTIAL).
FT      77      3615      BY SIMILARITY.
FT      3616      212      BY SIMILARITY.
FT      213      225      BY SIMILARITY.
FT      226      234      BY SIMILARITY.
FT      235      297      BY SIMILARITY.
FT      298      310      BY SIMILARITY.
FT      311      319      BY SIMILARITY.
FT      320      337      BY SIMILARITY.
FT      338      350      BY SIMILARITY.
FT      351      359      BY SIMILARITY.
FT      360      381      BY SIMILARITY.
FT      382      394      BY SIMILARITY.
FT      395      394      BY SIMILARITY.
FT      396      403      BY SIMILARITY.
FT      404      479      BY SIMILARITY.
FT      480      773      BY SIMILARITY.
FT      774      780      BY SIMILARITY.
FT      781      792      BY SIMILARITY.
FT      793      811      BY SIMILARITY.
FT      812      829      BY SIMILARITY.
FT      830      839      BY SIMILARITY.
FT      840      851      BY SIMILARITY.
FT      852      869      BY SIMILARITY.
FT      870      1168      BY SIMILARITY.
FT      1169      1175      BY SIMILARITY.
FT      1176      1187      BY SIMILARITY.
FT      1188      1206      BY SIMILARITY.
FT      1207      1224      BY SIMILARITY.
FT      1225      1234      BY SIMILARITY.
FT      1235      1246      BY SIMILARITY.
FT      1247      1263      BY SIMILARITY.
FT      1264      1287      BY SIMILARITY.
FT      1288      1287      BY SIMILARITY.

```

FT	DISULFID	1277	1293	BY SIMILARITY.
FT	DISULFID	1295	1304	BY SIMILARITY.
FT	DISULFID	1307	1322	BY SIMILARITY.
FT	DISULFID	1563	1572	BY SIMILARITY.
FT	DISULFID	1565	1579	BY SIMILARITY.
FT	DISULFID	1582	1591	BY SIMILARITY.
FT	DISULFID	1594	1610	BY SIMILARITY.
FT	DISULFID	1613	1628	BY SIMILARITY.
FT	DISULFID	1615	1638	BY SIMILARITY.
FT	DISULFID	1641	1650	BY SIMILARITY.
FT	DISULFID	1653	1668	BY SIMILARITY.
FT	DISULFID	1792	1839	BY SIMILARITY.
FT	DISULFID	1886	1932	BY SIMILARITY.
FT	DISULFID	1976	2021	BY SIMILARITY.
FT	DISULFID	2073	2218	BY SIMILARITY.
FT	DISULFID	2170	2215	BY SIMILARITY.
FT	DISULFID	2268	2313	BY SIMILARITY.

```

DR PROSITE; PS00290; IG_MHC; 1.
KW Antigen; Signal; Immunoglobulin domain.
FT SIGNAL 1 37 POTENTIAL.
FT CHAIN 38 213 IMMUNOGLOBULIN LAMBDA-LIKE POLYPEPTIDE 1.
FT DOMAIN 97 108 J REGION (BY SIMILARITY TO LAMBDA
FT DOMAIN 109 213 LIGHT-CHAIN).
FT DOMAIN 213 213 C REGION (BY SIMILARITY TO LAMBDA
FT SEQUENCE 213 AA; 22963 MW; 9133A742B943C79 CRC64;
SQ
Query Match 6.7%; Score 162.5; DB 1; Length 213;
Best Local Similarity 30.5%; Pred. No. 0.0003;
Matches 36; Conservative 21; Mismatches 44; Indels 17; Gaps 3;

OY 327 QPR-EPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDLAVEMSNQGP-ENNYKTPPYLTD 384
DB 109 QPKATPSTLTLPSPSEELQANKATLVCLMNDYFPIILVTWKADSTPTTGQVEMTTPSKQ 168
OY 385 SDGSFFLYSKLTVDKSRWQGNVFSQSVMEHALNHHYTKSLSPGLQDDETCANAO 442
DB 169 SNNKXAASYSYLTLPTEQWRSRSYSQYVME-----GSTVETKTAAPAE 211

RESULT 72
LAC2_MOUSE STANDARD; PRT; 104 AA.
AC P01844;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-2 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83014953; PubMed=6812053;
RA Sealing E., Miller J., Wilson R., Storch U.;
RT "Evolution of mouse immunoglobulin lambda genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=82274221; PubMed=6287422;
RA Mu G.E., Govindji N., Hozumi N., Murialdo H.;
RT "Nucleotide sequence of a chromosomal rearranged lambda 2
RL Nucleic Acids Res. 10:3831-3843(1982).
RN [3]
RP SEQUENCE FROM N.A. (MOPC 315).
RX MEDLINE=82220143; PubMed=6283385;
RA Bothwell A.L.M., Paekind M., Rech M., Imanishi-Kari T., Rajewsky K.,
RT "Somatic variants of murine immunoglobulin lambda light chains.";
RL Nature 298:380-382(1982).
RN [4]
RP SEQUENCE (MOPC 315).
RX MEDLINE=74048693; PubMed=4760498;
RA Dugan E.S., Bradshaw R.A., Simms E.S., Eisen H.N.;
RT "Amino acid sequence of the light chain of a mouse myeloma protein
RL (MOPC-315).";
RN [5]
RP SEQUENCE (MOPC 315).
RX MEDLINE=8123782; PubMed=6165998;
RA Azuma T., Steiner L.A., Eisen H.N.;
RT "Identification of a third type of lambda light chain in mouse
RL immunoglobulins.";
CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ch).
CC -----
DR EMBL; J00595; AAA39151.1; -.
DR PIR; C93922; L2MS.
DR HSSP; P01844; 2MCG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 6 99 IG-LIKE.
FT DISULFID 27 85
FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
FT SEQUENCE 104 AA; 11254 MW; CE4B67B868862D3 CRC64;
SQ
Query Match 6.7%; Score 161; DB 1; Length 104;
Best Local Similarity 34.3%; Pred. No. 0.0016;
Matches 35; Conservative 15; Mismatches 48; Indels 4; Gaps 2;

OY 327 QPR-EPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDLAVEMSNQGP-ENNYKTPPYLDS 385
DB 1 QPKATPSTLTLPSPSEELQANKATLVCLISNFSPTGVTAMKANGTPTTGQVDSNPTKE 60
OY 386 DGSEFFLYSKLTVDKSRWQGNVFSQSVMEHALNHHYTKSLS 427
DB 61 GNKTMASFLHITSDDQWRSNHSFTCYVTHE---GDTVEKSLDS 99

RESULT 73
HB2D_PIG STANDARD; PRT; 258 AA.
ID HB2D_PIG
AC P15983;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SLA class II histocompatibility antigen, DQ haplotype D beta chain
DE precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90361905; PubMed=2391424;
RA Gustafsson K., Leguern C., Hirsch F., Germana S., Pratt K.,
RT "Class II genes of miniature swine. IV. Characterization and
RT expression of two allelic class II DQB cDNA clones.";
RL J. Immunol. 145:1946-1951(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ch).
CC -----
DR EMBL; M31498; AAA31085.1; -.
DR HSSP; P13760; 2SEB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig_1.

```

DR Pfam: PF00969; MHC II beta; 1.
DR Prodom: PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 31
FT CHAIN 32 258
FT 32 258 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
DQ HAPLOTYPE D BETA CHAIN.
FT DOMAIN 32 123
FT DOMAIN 124 217 EXTRACELLULAR BETA-1.
FT DOMAIN 218 227 EXTRACELLULAR BETA-2.
FT TRANSMEM 228 248 CONNECTING PEPTIDE.
FT DOMAIN 249 258
FT DISULFID 44 108 CYTOPLASMIC TAIL.
FT DISULFID 146 202 BY SIMILARITY.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 258 AA; 29262 MW; E3AC75110AED47C3 CRC64;

Query Match 6.7%; Score 161; DB 1; Length 258;
Best Local Similarity 28.4%; Pred. No. 0.00048;
Matches 46; Conservative 28; Mismatches 80; Indels 8; Gaps 4;

Qy 262 WYVDGVEVNAKTPREQVNSTYRVSVLYTLHODMLNGKVKVSKALPAPIEKTI 321
54 WSDVDR-IYNQGEFLRFSDMGEYRAVTLGRPDADYLGQKALEQKRAELDYCKANY 112
Db 322 STAKG--QPREQVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMNSQOPEN-NYK 377
113 QIEGGTTLQRRVQPTVTISPSKAEALNHNHLLVCAVTDYFPQVKQVFRNGQEBYAGV 172
Qy 378 TTPPVLDSDGSFFLYSKLTVDSKRMQGVNFGSVNHEALN 419
Db 173 STPLIRNGD--WTYQVLVLEMLNLRGQVYTCREHSSLSQS 211

RESULT 74
CAM1_BRARE STANDARD; PRT; 1197 AA.
ID CAM1_BRARE
AC Q90478;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule LI.1 (N-CAM LI.1) (Fragment).
GN NADL1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OK NCBI_TaxID=7955;
RN (1)
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC TISSUE=embryo;
RC MEDLINE=96155762; PubMed=8568941;
RA Tongiorgi E., Bernhard R.R., Schachner M.;
RA "Zebrafish neurons express two LI-related molecules during early
RT axonogenesis.";
RL J. Neurosci. Res. 42:547-561(1995).
CC -1- FUNCTION: Cell adhesion molecule with an important role in the
CC development of the nervous system. Involved in neuron-neuron
CC adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
CC to axonin on neurons (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in postmitotic neurons in 16-36
CC hour embryos, including those in the brain, cranial ganglia and
CC olfactory placodes, and in all classes of spinal
CC neurons.
CC -1- DEVELOPMENTAL STAGE: Onset of expression correlates with the
CC initiation of axonogenesis in 16-36 hour embryos.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC LI/neurofascin/NCAM family.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89204; CAA61490.1; .
DR PIR; T30581; T30581.
DR HSSP; P20241; ICFB.
DR ZFIN; ZDB-GENE-980526-512; naatl.1.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FnIII subd.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; ig; 6.
DR PRINTS; PR00014; FNTPBIII.
DR SMART; SM00060; FN3; 5.
DR PROSITE; PS50835; IG_LIKE; 6.
DR Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;
KW Transmembrane; Repeat; Immunoglobulin domain.
FT NON_TER 1
FT DOMAIN <1 1054
FT TRANSMEM 1055 1075
FT DOMAIN 1076 1197
FT DOMAIN <1 58
FT DOMAIN 69 160
FT DOMAIN 165 263
FT DOMAIN 268 355
FT DOMAIN 360 442
FT DOMAIN 451 541
FT DOMAIN 546 638
FT DOMAIN 645 739
FT DOMAIN 744 849
FT DOMAIN 850 948
FT DOMAIN 952 1029
FT DISULFID 92 143
FT DISULFID 199 247
FT DISULFID 289 339
FT DISULFID 383 432
FT DISULFID 472 525
FT CARBOHYD 135 135
FT CARBOHYD 149 149
FT CARBOHYD 221 221
FT CARBOHYD 298 298
FT CARBOHYD 414 414
FT CARBOHYD 421 421
FT CARBOHYD 438 438
FT CARBOHYD 449 449
FT CARBOHYD 708 708
FT CARBOHYD 959 959
FT CARBOHYD 968 968
FT CARBOHYD 1002 1002
FT CARBOHYD 1027 1027
SQ SEQUENCE 1197 AA; 132860 MW; 7CE1509EBFAC7B28 CRC64;

Query Match 6.6%; Score 159.5; DB 1; Length 1197;
Best Local Similarity 19.3%; Pred. No. 0.0042;
Matches 94; Conservative 80; Mismatches 186; Indels 127; Gaps 22;

Qy 9 HLLVLTGLALPAATQGNKVVLTCTASQKKSIOFPHKNSNOKIIGNGSF 68
Db 259 HTYVTEAA--PYWTRSEEHLYAPGETVRLDCKADGIPANITW-SINGVPVSGTD--- 312
Qy 69 LTKGPSKLNDRADSRRLSDGNFPLITKNLKIEDSDTYICEVDEQKEEVOLLVGLTAN 128
Db 313 -----VDPRRRV---SSGKLILSNVEFSPTAVYQGEAVNKGISILI-----N 351

```

QY 129 SDTHLLQSGSLTLT-----LESPPSSPSVOCRS-----PRGNKIQGGK 167
DB 352 THVHVVELPAQLITPTDERLYQATAGATWLDORTGSLPKHWEILSLALSNAKISQ 411
QY 168 T---LSVSOLELQDSGTWTCTVTLONQKKVEFKDIVCPAPRPPKSCKTHTCPELLGCP 223
DB 412 TTNGSLKISINSESDSNRYTCSVSETNKSISADVEVL-----NRTK-----IVGP 456
QY 224 SVFLRPPPKDTLMTSRTPVTCVVDVSHDEPEKFMVYGVGVHNAKTKPREQVNS 283
DB 457 PONTLVIVIGSDAILCK-----YTVDNHLSPTVQ--WNKDG--KHITSTNSDDKTHE 505
QY 284 TYRVVSVLTTLVHQDLNGEKYCKVSNKALPAPIEKTSKAKGQPREQVYTLPPSRDEL 343
DB 506 IEGSLKVLVDVQWEDM---GIYSCVEST--TLDSDTASGYITVQDKPDPQSLKLSKMER- 560
QY 344 TKQNVSLTCLVKGFPSPDIAYVESNQGPENNYKTPPYLD-----SDGSFFLYSKLT 396
DB 561 -----SVTISWMPSS--VENNSPVTREVIEKNNEGTPDEGQWQKYSVS 601
QY 397 VDKSRWQGVNPSGVHMEALH-----NHYSKLSLSPQLQDETCAEAQDELDELGL 449
DB 602 QDIDWR--SICSYSKYHFOIRAVNSIGTSAPTESSLSYS-----TPAAKPDTPNPNV 652
QY 450 WT--TDP 454
DB 653 MTLSTDP 659

```

RESULT 75

LAC_HUMAN STANDARD; PRT; 105 AA.

```

AC PO1842; P80423;
DT 21-JUL-1996 (Rel. 01, Created)
DT 21-JUL-1996 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig lambda chain C regions.
GN IGLC1 AND IGLC2 AND IGLC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_Taxid=9606;
RN [1]
RP SEQUENCE (BENCE-JONES PROTEIN SH).
RX MEDLINE=70166723; PubMed=4909564;
RA Titani K., Wikler M., Shinoda T., Putnam F.W.;
RT "The amino acid sequence of a lambda type Bence-Jones protein. 3. The
RT complete amino acid sequence and the location of the disulfide
RT bridges."
RL J. Biol. Chem. 245:2171-2176(1970).
RN [2]
RP SEQUENCE (BENCE-JONES PROTEIN X), AND DISULFIDE BONDS.
RX MEDLINE=69088380; PubMed=4883841;
RA Milstein C., Clegg J.B., Jarvis J.M.;
RT "Immunoglobulin lambda-chains. The complete amino acid sequence of a
RT Bence-Jones protein."
RL Biochem. J. 110:631-652(1968).
RN [3]
RP SEQUENCE (BENCE-JONES PROTEIN NIG-64).
RX MEDLINE=83186114; PubMed=6404900;
RA Kametani F., Takayasu T., Suzuki S., Shinoda T., Okuyama T.,
RA Shimizu A.;
RT "Comparative studies on the structure of the light chains of human
RT immunoglobulins. IV. Assignment of a subgroup."
RL J. Biochem. 93:421-429(1983).
RN [4]
RP SEQUENCE (BENCE-JONES PROTEIN KERN).
RX MEDLINE=71150336; PubMed=5549568;
RA Ponstingl H., Hees M., Hilschmann N.;
RT "Structural rule of antibodies. Primary structure of a monoclonal
RT immunoglobulin-L-chain of the lambda type, subgroup IV (Bence-Jones-
RT protein Kern). V. The complete amino acid sequence and its genetic
RT interpretation."

```

```

RL Hoppe-Seyler's Z. Physiol. Chem. 352:247-266(1971).
RN [5]
RP PARTIAL SEQUENCE (MYELOMA PROTEIN NEMM).
RX MEDLINE=74109253; PubMed=4814727;
RA Chen B.L., Poljak R.J.;
RT "Amino acid sequence of the (lambda) light chain of a human myeloma
RT immunoglobulin (IgG New)."
RL Biochemistry 13:1295-1302(1974).
RN [6]
RP SEQUENCE (DOT).
RX MEDLINE=95255298; PubMed=7737190;
RA Stopini M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;
RT "Characterization of the two unique human anti-flavon monoclinal
RT immunoglobulins."
RL Eur. J. Biochem. 228:886-893(1995).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF NEMM.
RX MEDLINE=75046825; PubMed=4215080;
RA Poljak R.J., Amzel L.M., Avey H.P., Chen B.L., Phizackerley R.P.,
RA Saul F.;
RT "The three-dimensional structure of the fab' fragment of a human
RT myeloma immunoglobulin at 2.0-A resolution."
RL Proc. Natl. Acad. Sci. U.S.A. 71:3440-3444(1974).
RN [8]
RP PARTIAL SEQUENCE (BENCE-JONES PROTEIN MCG).
RX MEDLINE=75013804; PubMed=4415202;
RA Felt J.W., Deutsch H.F.;
RT "Primary structure of the Mcg lambda chain."
RL Biochemistry 13:4102-4114(1974).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF MCG.
RX Edmundson A.B., Ely K.R., Abola E.E., Schiffer M.,
RA Panagiotopoulos N.;
RT "Rotational allomerism and divergent evolution of domains in
RT immunoglobulin light chains."
RL Biochemistry 14:3953-3961(1975).
RN [10]
RP X-RAY CRYSTALLOGRAPHY OF MCG.
RX MEDLINE=90133913; PubMed=2515285;
RA Ely K.R., Herron J.N., Harker M., Edmundson A.B.;
RT "Three-dimensional structure of a light chain dimer crystallized in
RT water. Conformational flexibility of a molecule in two crystal
RT forms."
RL J. Mol. Biol. 210:601-615(1989).
RN [11]
RP SEQUENCE FROM N.A.
RX MEDLINE=82080680; PubMed=6273747;
RA Hicter P.A., Hollis G.F., Korsmeyer S.J., Waldmann T.A., Leder P.;
RT "Clustered arrangement of immunoglobulin lambda constant region genes
RT in man."
RL Nature 294:536-540(1981).
RN [12]
RP MISCELLANEOUS: The sequence shown is the Kern/Oz-/MCG- chain
RP found in proteins SH, X, and NIG-64. The Kern protein has the
RP Kern+ marker, the NEMM protein has the Oz+ marker, the MCG protein
RP has the Kern+ marker, and the MCG+ marker.
RN [13]
RP MISCELLANEOUS: Six tandem lambda-type genes were identified and
RP the 3 most 5' were sequenced. These correspond to the MCG sequence
RP (lambda-1), the Kern/Oz- sequence (lambda-2) and the Kern/Oz+
RP sequence (lambda-3).
RN [14]
RP SIMILARITY: Contains 1 immunoglobulin-like domain.
RN [15]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@isb-sib.ch).
RN [16]
RP EMBL: J00253; AA59107.1; -
RP EMBL: L38562; AAB36581.1; ALT_INIT.
RP EMBL: X51754; CAA36569.1; ALT_INIT.
RP EMBL: X51755; CAA36049.1; -

```

```
DR EMBL: X51755; CAA36051.1; -
DR PIR: A92057; L2HU-92.
DR PDB: 2MCG; 15-JUL-92.
DR PDB: 7FAB; 31-JAN-94.
DR PDB: 1AOK; 04-FEB-98.
DR PDB: 1LIL; 15-MAY-97.
DR Genew; HGNC:5855; IGLC1.
DR Genew; HGNC:5856; IGLC2.
DR Genew; HGNC:5857; IGLC3.
DR MIM; 147220; -.
DR GO: 0003823; P: antigen binding; NAS.
DR GO: 0006955; P: immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00407; IGL1; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain, Immunoglobulin C region; Bence-Jones protein;
3D-structure.
FT NON_TER 1 1
FT DOMAIN 6 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104
FT VARIANT 5 5 A->N (IN MCG+ MARKER).
FT VARIANT 7 7 /FTID=VAR_003898.
FT VARIANT 7 7 S->T (IN MCG+ MARKER).
FT VARIANT 45 45 /FTID=VAR_003899.
FT VARIANT 45 45 S->G (IN KERN+ MARKER).
FT VARIANT 56 56 /FTID=VAR_003900.
FT VARIANT 82 82 T->K (IN MCG+ MARKER).
FT VARIANT 82 82 /FTID=VAR_003901.
FT VARIANT 82 82 R->K (IN OZ+ MARKER).
FT VARIANT 82 82 /FTID=VAR_003902.
FT STRAND 8 11
FT HELIX 15 19
FT TURN 20 21
FT STRAND 24 32
FT STRAND 38 43
FT TURN 44 45
FT STRAND 46 48
FT STRAND 52 59
FT STRAND 61 62
FT STRAND 65 72
FT STRAND 75 80
FT HELIX 84 89
FT STRAND 94 99
SQ SEQUENCE 105 AA; 11236 MW; DCD9C7C201C13CC2 CRC64;

Query Match 6.6%; Score 159; DB 1; Length 105;
Best Local Similarity 34.1%; Pred. No. 0.00021;
Matches 31; Conservative 18; Mismatches 40; Indels 2; Gaps 2;

QY 327 GQR-EEQVYTLPSRDELTKQVSLTCLYKGFPSDIAIENWESNGQP-ENNYKTTTPVVD 384
DB 1 QPAAASVTLFPSSSELDANKATLVCLISDFPGAIVTAMKADSSPVKAGVETTPSKQ 60
QY 385 SDGSFFLYSKLTVDKRMQGNVFCSCVWHE 415
DB 61 SNKKVAASSYLTLPQGMKSHRSYSQGVTHE 91

RESULT 76
ID _VCAM1_HUMAN STANDARD; PRT; 739 AA.
AC P19330;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1) (CD106 antigen)
DE (INCM-100).
GN VCAM1 OR L1CAM.
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=umbilical vein;
RX MEDLINE=91016951; PubMed=1699207;
RT Polte T., Newman W., Gopal T.V.;
RL "Full length vascular cell adhesion molecule 1 (VCAM-1).";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90090619; PubMed=2688898;
OSborn L., Hession C., Tizard R., Vassallo C., Luhnowskyj S.,
RA Chi-Rosso G., Lobb R.;
RT "Direct expression cloning of vascular cell adhesion molecule 1, a
cytokine-induced endothelial protein that binds to lymphocytes.";
RL Cell 59:1203-1211(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91352090; PubMed=1715583;
CYbulsky M.I., Fries J.W.U., Williams A.J., Sultan P., Eddy R.,
RA Byers M., Shows T., Gimbrone M.A. Jr., Collins T.;
RT "Gene structure, chromosomal location, and basis for alternative mRNA
splicing of the human VCAM1 gene.";
Proc. Natl. Acad. Sci. U.S.A. 88:7859-7863(1991).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91201302; PubMed=1707873;
RA Hession C., Tizard R., Vassallo C., Schiffer S.B., Goff D., Moy P.,
RA Chi-Rosso G., Luhnowskyj S., Lobb R., Osborn L.;
RT "Cloning of an alternate form of vascular cell adhesion molecule-1
(VCAM1).";
J. Biol. Chem. 266:6682-6685(1991).
RN [5]
RP SEQUENCE FROM N.A., AND VARIANTS PHE-318; ALA-384; ALA-413 AND
LEU-716.
RA Rieder M.J., Armet T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA Rajkumar N.R., Toch E.J., Yi Q., Nickerson D.A.;
RL Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Retinal pigment epithelium;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinchi P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Wuzny D.M., Sodergren E.U., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.T., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.
RX MEDLINE=95147978; PubMed=7531291;
RA Jones E.Y., Harlos K., Bottomley M.J., Robinson R.C., Driscoll P.C.,
RA Edwards R.M., Clements J.M., Dudgeon T.J., Stuart D.I.;
RT "Crystal structure of an integrin-binding fragment of vascular cell
adhesion molecule-1 at 1.8-A resolution.";
Nature 373:539-544(1995).
```

```

RM [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RX MEDLINE=95296382; Pubmed=7539925;
RA Wang J.-H., Pepinsky R.B., Stehle T., Liu J.-H., Karpusas M.,
RA Browning B., Osborn L.;
RT "The crystal structure of an N-terminal two-domain fragment of
RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on
RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin
RT interaction."
RL Proc. Natl. Acad. Sci. U.S.A. 92:5714-5718(1995).
RM [9]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RX Wang J.-H., Stehle T., Pepinsky R.B., Liu J.-H., Karpusas M.,
RA Osborn L.;
RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A
RT resolution."
RL Acta Crystallogr. D 52:369-379(1996).
CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VIA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=Long;
CC IsoId=P19320-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P19320-2; Sequence=VSP_002580;
CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
CC well as on macrophage-like and dendritic cell types in both normal
CC and inflamed tissue.
CC -1- INDUCTION: By cytokines (e.g. IL-1, TNF-alpha).
CC -1- PTM: Sialoglycoprotein.
CC -1- DISEASE: May play an important role in the genesis of
CC atherosclerosis and rheumatoid arthritis.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD106 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X53051; CA37218.1; -.
DR EMBL; M30257; AAA51917.1; ALT_TERM.
DR EMBL; M73255; AAA61270.1; -.
DR EMBL; M60335; AAA61269.1; -.
DR EMBL; AF536818; AAM96190.1; -.
DR EMBL; BC017276; AAH17276.1; -.
DR PIR; A41288; A41288.
DR PIR; B41288; B41288.
DR PDB; 1YCA; 15-SEP-95.
DR PDB; 1VSC; 20-JUN-96.
DR PDB; 1IJ9; 07-NOV-01.
DR Genew; HGNC:12663; VCAM1.
DR MIM; 192225; -.
DR InterPro; IPR003987; ICAM VCAM-1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig; 6.
DR PRINTS; PR01472; ICAMVCAM1.
DR PRINTS; PR01474; VCAM1.
DR SMART; SM00408; IGC2_3.
DR PROSITE; PS00835; IG_LIKE; 5.
KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;

```

```

KW Repeat; Signal; Alternative splicing; Polymorphism; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 739
FT DOMAIN 25 698
FT TRANSMEM 699 720
FT DOMAIN 721 739
FT DOMAIN 25 105
FT DOMAIN 109 212
FT DOMAIN 223 309
FT DOMAIN 312 399
FT DOMAIN 408 506
FT DOMAIN 511 595
FT DOMAIN 600 684
FT DISULFID 47 95
FT DISULFID 52 99
FT DISULFID 137 195
FT CARBOHYD 273 273
FT CARBOHYD 365 365
FT CARBOHYD 417 417
FT CARBOHYD 463 531
FT CARBOHYD 561 561
FT VARSPIC 310 402
FT VARIANT 318 318
FT VARIANT 384 384
FT VARIANT 413 413
FT VARIANT 716 716
FT STRAND 26 30
FT STRAND 34 38
FT TURN 39 40
FT STRAND 43 50
FT STRAND 56 61
FT TURN 62 63
FT STRAND 70 74
FT TURN 75 76
FT STRAND 77 82
FT HELIX 87 89
FT STRAND 92 99
FT TURN 100 101
FT STRAND 102 114
FT STRAND 120 123
Query Match 6.5%; Score 157; DB 1; Length 739;
Best local similarity 20.6%; Pred. No. 0.0034;
Matches 95; Conservative 70; Mismatches 139; Indels 158; Gaps 21;
16 LALLPATQGNKV-----VLGKGDVVELTCAQSQKSIQPHMKNSNQIKLG---NQ 65
15 LMTFPAASQAFKLETTPEBSRYLQIGDSVSLTCTTCESPFPSWRQIDSPLENGKVTNE 74
66 G-SFLTKGPSKLNDR-----ADSRRLMDQG---NFP----- 93
75 GTTSTLTMNPVVSFENHSYLTCTACESRK--LEKGIVETIYSPFKDEIHLSGLPBAKGP 132
94 -----LIINKLK-----IEDSD-----TYICEVED----- 113
133 ITVKCSVADVYPPDRLEIDILKGDHLMKSOEFLEDADRKSLKTSLEVTFTPVIEDIGKV 192
114 -----QKEBYOLLVVG-----LTANSDDTHLQGSLLTTLTLESPGGS 149
193 LVCRAKLHIDEMDSVPTVROAVKELQVISPKNTVISVSTKLQDSGVMTTSSSGLP 252
150 SPSPV-----QCRSPRGNIOGAKTISVQLBDSGTWTC-----TVLQNOKKVEFKIDIVP 201
253 APEIFMSKLDNGMLQHLSGNATITTLAMRMEDSGIIVCGVNLIGNRRVELIIVG--- 309
202 CPADPEPSCKTHTCPPLLGGPSVFLFPKPKDTLMISRPDEVTCVVVDVSHEDPEVKEN 261

```

DB 310 -----EK-----PFTVEISPPRIAGIGSVMLTCSWGC--ESPFSWR 348
QY 262 WYDGVGVNNAKTRPEEQVNSTYRVVSLTVLHQMNGKEYCKVNS--NKALPPIRX 319
DB 349 TQIDS-----PLSGKVSSEGTNST-----LTPSVSEFHEHSYLCVTGCKHKLKNGIQV 398
QY 320 TISKAKQPREPOVYLTLPSPRDELTKNOVSLTCLVGVGFPSD 361
DB 399 ELVXF---PRDPEI---EMSGGLVNSVTVSCKVPSTVPLD 434

RESULT 77
SHS1 HUMAN STANDARD: PRT: 503 AA.
AC F78324; Q00683; Q43799; Q8N517; Q8TAL6; Q9H022; Q9UDX2; Q9UIJ6;
AC Q9Y409;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor
DE (SHP substrate-1) (SHPS-1) (Inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (S1rp-alpha-1) (S1rp-alpha-
DE 3) (MYD-1 antigen) (Brain Ig-like molecule with tyrosine-based
DE activation motifs) (Bic) (Macrophage fusion receptor) (p84).
GN PTPN1 OR SHPS1 OR S1RP OR MYD1 OR BIT OR MFR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=97223399; PubMed=9070220;
RA Yamao T., Matosaki T., Amano K., Matsuda Y., Takahashi N., Ochi F.,
RA Fujioke Y., Kasuga M.;
RT "Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal
RT localization of genes";
RL Biochem. Biophys. Res. Commun. 231:61-67(1997).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 1), POLYMORPHISM, PHOSPHORYLATION,
RP N-GLYCOSYLATION, AND INTERACTIONS WITH PTPN11, PTPN6 AND GRB2.
RC TISSUE=Placenta;
RX MEDLINE=97215901; PubMed=9062191;
RA Khatichenkov A., Chen Z., Sures I., Wang H., Schilling J.,
RA Ullrich A.;
RT "A family of proteins that inhibit signalling through tyrosine kinase
RT receptors";
RL Nature 386:181-186(1997).
RN (3)
RP SEQUENCE FROM N.A. (ISOFORM 3), AND VARIANTS ASP-95; LEU-96; ASN-100;
RP ARG-107; GLY-109 AND VAL-131.
RC TISSUE=Monocytes;
RX MEDLINE=98143722; PubMed=9485180;
RA Brooke G.P., Parsons K.R., Howard C.J.;
RT "Cloning of two members of the S1rp alpha family of protein tyrosine
RT phosphatase binding proteins in cattle that are expressed on monocytes
RT and a subpopulation of dendritic cells and which mediate binding to
RT CD4 T cells";
RL Eur. J. Immunol. 28:1-11(1998).
RN (4)
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS LEU-44; THR-50; THR-52;
RP ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107; GLY-109 AND
RP VAL-131.
RC TISSUE=Brain;
RX MEDLINE=20053880; PubMed=10585853;
RA Sano S.-I., Ohnishi H., Kubota M.;
RT "Gene structure of mouse BIT/SHPS-1";
RL Biochem. J. 344:667-675(1999).
RN (5)
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
BA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
BA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
BA Buck D., Burrill M.D., Butler A.P., Carder C., Carter N.P.,
BA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
BA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
BA Coulson A., Coville G.J., Dearden R., Dhani P.D., Dunn M.,
BA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
BA Gething D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
BA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
BA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
BA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
BA Levasaisho M.H., Leverha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
BA Marsh V.L., Martin S.L., McConachie L.J., McIay K., Murray A.A.,
BA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
BA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
BA Phillimore B.J.C.T., Prichalingam S.R., Plumb R.W., Ramsey H.,
BA Rice C.M., Ross M.T., Scott C.E., Serna H.K., Showkeen R., Sims S.,
BA Skues C.D., Smith M.L., Soderlund C., Stewart C.A., Sultson J.E.,
BA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
BA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
BA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Bentley D.R., Beck S.,
BA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
BA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20";
RL Nature 414:865-871(2001).
RN (6)
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS LEU-44; THR-50;
RP THR-52; ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107;
RP GLY-109 AND VAL-131.
RC TISSUE=Brain, Kidney, and Skin;
RX MEDLINE=22388257; PubMed=1477932;
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.T., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson W.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN (7)
RP FUNCTION, AND INTERACTIONS WITH FYB, SCAIP2 AND PTK2B.
RX MEDLINE=99401000; PubMed=10469599;
RA Timms J.F., Swanson K.D., Marie-Cardine A., Raab M., Rudd C.E.,
RA Schraven B., Neel B.G.;
RT "SHPS-1 is a scaffold for assembling distinct adhesion-regulated
RT multi-protein complexes in macrophages";
RL Curr. Biol. 9:927-930(1999).
RN (8)
RP PHOSPHORYLATION BY JAK2, AND INTERACTIONS WITH PTPN11 AND JAK2.
RX MEDLINE=20428742; PubMed=10842184;
RA Stofega M.R., Argenteinger L.S., Wang H., Ullrich A., Carter-Su C.;
RT "Negative regulation of growth hormone receptor/JAK2 signaling by
RT signal regulatory protein alpha";
RL J. Biol. Chem. 275:28222-28229(2000).
RN (9)
RP FUNCTION, AND INTERACTION WITH CD47.
RX MEDLINE=21400825; PubMed=11509594;
RA Latour S., Tanaka H., Demeure C., Mateo V., Rubio M., Brown E.J.,
RA Maliszewski C., Lindberg P.P., Oldenborg A., Ullrich A.,
RA Delapese G., Sarfati M.;
RT "Bidirectional negative regulation of human T and dendritic cells by

RT CD47 and its cognate receptor signal-regulator protein- α :
 RT down-regulation of IL-12 responsiveness and inhibition of dendritic
 cell activation";
 RL J. Immunol. 167:2547-2554(2001).
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
 CC as docking protein and induces translocation of PTPN6, PTPN11 and
 CC other binding partners from the cytosol to the plasma membrane.
 CC Supports adhesion of cerebellar neurons, neurite outgrowth and
 CC glial cell attachment. May play a key role in intracellular
 CC signaling during synaptogenesis and in synaptic function (By
 CC similarity). Involved in the negative regulation of receptor
 CC tyrosine kinase-coupled cellular responses induced by cell
 CC adhesion, growth factors or insulin. Mediates negative regulation
 CC of phagocytosis, mast cell activation and dendritic cell
 CC activation. CD47 binding prevents maturation of immature dendritic
 CC cells and inhibits cytokine production by mature dendritic cells.
 CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
 CC macrophages, where it primarily binds PTPN6. Binds GRB2 in vitro.
 CC Binds FGR (By similarity). Binds JAK2 irrespective of its
 CC phosphorylation status and forms a stable complex. Binds SCAP1
 CC and/or SCAP2. The resulting complex recruits Fyb. Binds PTK2B.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=P78324-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P78324-2; Sequence=VSP_007030;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=P78324-3; Sequence=VSP_007029;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain.
 CC Detected on myeloid cells, but not T cells. Detected at lower
 CC levels in heart, placenta, lung, testis, ovary, colon, liver,
 CC small intestine, prostate, spleen, kidney, skeletal muscle and
 CC pancreas.
 CC -1- PTM: N-glycosylated.
 CC -1- PTM: Phosphorylated on tyrosine residues in response to
 CC stimulation with EGF, growth hormone, insulin and PDGF.
 CC Dephosphorylated by PTPN11.
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D86043; BAA12974.1; -;
 DR EMBL: Y10375; CAA71403.1; -;
 DR EMBL: Y11047; CAA71944.1; ALT_INIT.
 DR EMBL: AB023430; BAA87929.1; -;
 DR EMBL: AC004833; AAF19260.1; -;
 DR EMBL: AL034563; CAB38874.1; -;
 DR EMBL: AL049634; CAB46662.1; ALT_SEQ.
 DR EMBL: AL117335; CAC12723.1; -;
 DR EMBL: BC026692; AAH26692.1; -;
 DR EMBL: BC033093; AAH33093.1; -;
 DR EMBL: BC038510; AAH38510.1; -;
 DR PIR: JCS287; JCS287
 DR GeneW: HGNC:9662; PTPN51.
 DR MIM: 602461; -;
 DR GO: GO:0005866; C:plasma membrane; TAS.
 Query Match 6.5%; Score 156.5; DB 1; Length 503;
 Best Local Similarity 24.8%; Pred. No. 0.0022;
 Matches 79; Conservative 45; Mismatches 153; Indels 41; Gaps 14;
 QY 136 GQSILTLDESPGSSPVQCRSPROKTIQGGKTLVSQLELDGSGTWTCTVL-QNOKVE 194

Db 73 GPANELYNQKEGHFPVATTVSESTKRENMDFSSISNITPADAGTYCYVFRKSPDTE 132
 QY 195 FKIDIVCPAPPEPSCCKTTCPELGGPSVFLFPPEPKOTLMSKTPETVCVVVDVSH 254
 Db 133 FK-----SGAGTELVSVAKESAP-VSGPARATP-----QHTVSEFCSESHGSPR 177
 QY 255 DPEVKFMVYDGVENVHAKTK--PRE-----QVNSTRVVSVLTVLHQDMLNGEKYCKV 308
 Db 178 D--ITLHMFKNGLNLSFQTNNDPVGSSVSISHTKVLTREDVASYI-----CEV 229
 QY 309 SNKALPA-PIEKT--ISKAGQPREPQVYTLPPSRDELTRKQVSLTLVKGFPESDI 365
 Db 230 AHVTLQDDPLRGFTANLSETRIVPPTLEV-TQGPVRAE---NQVWVTCQVRKFYQRLQLT 285
 QY 366 WESGQEPENNYKTPFPVLDDGSEFLYSKLTVDKSRMQGQNVFSCVME---ALAHNYT 422
 Db 286 WLENGVSRRETASTVTENKDGTYNMWSLLVNVSAHRDVKLTQVEHDOQPAVSKSHD 345
 QY 423 QKSLSPGLQDDETCAE 440
 Db 346 LK-VSAHPKEDGSGVTAHE 362
 RESULT 78
 LAC_PIG STANDARD; PRT; 105 AA.
 AC P01846;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig lambda chain C region.
 OS Sue scrofa (Pig).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 CC NCBI_TaxID=9823;
 RN [1]
 RP MEDLINE=78000254; PubMed=409425;
 RA Novotny J., Franek F., Margolies M.N., Haber E.;
 RT "Amino acid sequence of normal (microheterogeneous) porcine
 RT immunoglobulin lambda chains.";
 RL Biochemistry 16:3765-3772(1977).
 CC -1- MISCELLANEOUS: This chain was obtained from a mixture of normal
 CC immunoglobulins.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
 CC PIR: A02129; LIPG.
 DR HSSP: P01842; ZMCG.
 DR InterPro: IPR007110; Ig_1like.
 DR InterPro: IPR003597; Ig_C1.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig_1.
 DR SMART: SM00407; IgC1; 1.
 DR PROSITE: PS50835; IG_LIKE; 1.
 DR PROSITE: PS00290; IG_MHC; 1.
 KW Immunoglobulin domain; Immunoglobulin C region.
 FT NON TER 1
 FT DOMAIN 2 100 IG-LIKE.
 FT DISULFID 27 86
 FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
 SQ SEQUENCE 105 AA; 11003 MW; 3817AABD747C396 CRC64;
 Query Match 6.5%; Score 156; DB 1; Length 105;
 Best Local Similarity 35.2%; Pred. No. 0.00034;
 Matches 32; Conservative 15; Mismatches 42; Indels 2; Gaps 2;
 QY 327 QPR-EPQVYTLPPSRDELTRKQVSLTLVKGFPSPDI 384
 Db 1 QPKAPFVNLFPSSSELGNKATLVCLISDFRGATVVMKAGCTVTGQVETTKSKQ 60
 QY 385 SDGSFFLYSKLTVDKSRMQGQNVFSCVME 415
 Db 61 SNKTYAASSYLALASADMKSSSGFTCVTHE 91

```

RESULT 79
LACS_MUSSP STANDARD; PRT; 105 AA.
ID LACS_MUSSP
AC P20765;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-5 chain C region.
OS Mus spretus (Western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8196070; PubMed=3129289;
RA Mami F., Cazenave P.A., Kindt T.J.;
RT "Conservation of the immunoglobulin C lambda 5 gene in the Mus gene.";
RL EMO J. 7:117-122(1988).
CC -1 SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M3582; AAA39152.1; -.
CC DR HSSP; P01842; 2MCG.
CC DR InterPro; IPR007110; Ig-1like.
CC DR InterPro; IPR003597; IG_C1.
CC DR InterPro; IPR003006; IG_MHC.
CC DR Pfam; PF00047; Ig1; 1.
CC DR SMART; SM00407; Ig1; 1.
CC DR PROSITE; PS00835; IG_LIKE; 1.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR Immunoglobulin domain; Immunoglobulin C region.
CC KW NON_TER 1 100
CC FT DOMAIN 6 100
CC FT DISULFID 27 86
CC FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
CC SQ SEQUENCE 105 AA; 11674 MW; AAB417DF68471A17 CR664;

Query Match
Best Local Similarity 34.0%; Score 153.5; DB 1; Length 105;
Matches 35; Conservative 20; Mismatches 43; Indels 5; Gaps 3;

QY 327 QPR-EPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAYEWESNGQP-ENNYKTPPVLD 384
DB 1 QPKSPPLVTLFLPSKNIQANKVTTLVCLVSEFPGLVVDWVDGVPYVQGVETTPPSKQ 60
QY 385 SSGSEFFLVSKLTVDKSRMOQGNVFCSCVWHEALHNHYTKSLS 427
DB 61 TNNKTYWSSYLTLISQMPHRSYSCRYTHE--GNTVEKSVS 100

RESULT 80
LACS_RABIT STANDARD; PRT; 105 AA.
ID LACS_RABIT
AC P01847;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE.

```

```

RX MEDLINE=82091105; PubMed=6797414;
RA Garcia I., Jaron J.-C.;
RT "The primary structure of the constant region of Basilea-rabbit
RT immunoglobulin lambda-chains.";
RL Biochem. J. 197:177-183(1981).
CC -1 MISCELLANEOUS: This lambda chain expresses the c7 allotypic
CC specificity.
CC -1 SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC DR PIR; A02130; L7RB.
CC DR HSSP; P01842; 7FAB.
CC DR InterPro; IPR007110; Ig-1like.
CC DR InterPro; IPR003597; IG_C1.
CC DR InterPro; IPR003006; IG_MHC.
CC DR Pfam; PF00047; Ig1; 1.
CC DR SMART; SM00407; Ig1; 1.
CC DR PROSITE; PS00835; IG_LIKE; 1.
CC DR PROSITE; PS00290; IG_MHC; FALSE NEG.
CC KW Immunoglobulin domain; Immunoglobulin C region.
CC FT NON_TER 1 100
CC FT DOMAIN 6 100
CC FT DISULFID 27 86
CC FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
CC SQ SEQUENCE 105 AA; 11484 MW; B427513272E8663D CR664;

Query Match
Best Local Similarity 32.7%; Score 153; DB 1; Length 105;
Matches 32; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

QY 331 PQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAYEWESNGQP-ENNYKTPPVLDSDGSF 389
DB 6 PSVILFPPSSSEELKONKATVLCISDFPRVYKVMWKADGNSVTVGQVDTTPSKSNKY 65
QY 390 FLYSKLTVDKSRMOQGNVFCSCVWHEALHNHYTKSLS 427
DB 66 AASSFLHNTANQWXSQSVTCQVTHE--GHTVEKSLA 100

RESULT 81
LACS_RABIT STANDARD; PRT; 106 AA.
ID LACS_RABIT
AC P01839;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig kappa-B4 chain C region.
GN K-BAS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Basilea;
RA Heidmann O., Rougeon F.;
RT "Multispecificity of constant kappa light chain genes in the rabbit
RT genome: a b4b4 homozygous rabbit contains a kappa-bas gene.";
RL EMO J. 2:437-441(1983).
CC -1 MISCELLANEOUS: In Basilea rabbits, the major type of light chain
CC is lambda. The kappa chain shown is a minor component. All other
CC rabbit B allotypes have Cys-64.
CC -1 SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; V01241; CAA24558.1; -.
CC DR EMBL; V00885; -. NOT_ANNOTATED_CDS.
CC DR PIR; A02121; KARBS.

```

```
DR HSSE; P01842; 7FAB.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003597; Ig cl.  
DR InterPro; IPR003006; Ig_MHC.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00407; Igc1; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
DR PROSITE; PS00290; IG_MHC; FALSE NEG.  
KW Immunoglobulin domain; Immunoglobulin C region.  
FT NON TER 1  
FT DOMAIN 1 6 99 IG-LIKE.  
FT DISULFID 27 87  
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).  
SQ SEQUENCE 106 AA; 11279 MW; AFB928D8A853849 CRC64;  
  
Query Match 6.3%; Score 153; DB 1; Length 106;  
Best Local Similarity 33.7%; Pred. No. 0.00055;  
Matches 31; Conservative 19; Mismatches 36; Indels 6; Gaps 2;  
  
Qy 326 GQPREPQVYTLPPSDELTKNQVSLTCLVKGFYPSDIAVWESNGQP...NNYKTPP 381  
Db 1 GDPVAPSVLPPSPSEELITGTATIVCANKFPSDITVTKVDGTTQSGIENSKT--P 58  
  
Qy 382 VLSDSPFLLYSKLTVDKSRMOQGNVFCQSVW 413  
Db 59 QSPEDNTYLSSTSLTSLTSAQVNSHSVTCCEV 90  
  
RESULT 82  
DTC_HUMAN STANDARD; PRT; 383 AA.  
AC P01880;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig delta chain C region.  
GN IGHD.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OC NCBI_Taxid=9606;  
RN [1]  
RP SEQUENCE OF 1-162 (MYELOMA PROTEIN WAH).  
RX MEDLINE=82082419; PubMed=6947220;  
RA Putnam F.W., Takahashi N., Teraert D., Debuire B., Lin L.-C.;  
RT "Amino acid sequence of the first constant region domain and the  
RL hinge region of the delta heavy chain of human IgD";  
RN Proc. Natl. Acad. Sci. U.S.A. 78:6168-6172(1981).  
RN [2]  
RN SEQUENCE OF 158-383 (MYELOMA PROTEIN WAH).  
RX MEDLINE=81223768; PubMed=6787589;  
RA Lin L.-C., Putnam F.W.;  
RT "Primary structure of the Fc region of human immunoglobulin D:  
implications for evolutionary origin and biological function";  
RN Proc. Natl. Acad. Sci. U.S.A. 78:504-508(1981).  
RN [3]  
RN SEQUENCE OF 103-137 (MYELOMA PROTEIN NIG-65), AND CARBOHYDRATES.  
RX MEDLINE=82231401; PubMed=7092891;  
RA Takayasu T., Suzuki S., Kametani F., Takahashi N., Shinoda T.,  
Okuyama T., Kuneakata E.;  
RT "Amino acid sequence of galactosamine-containing glycopeptides in the  
RL hinge region of a human immunoglobulin D";  
RN Biochem. Biophys. Res. Commun. 105:1066-1071(1982).  
RN [4]  
RN SEQUENCE OF 158-383 (MYELOMA PROTEIN NIG-65).  
RX MEDLINE=8119406; PubMed=6785754;  
RA Shinoda T., Takahashi N., Takayasu T., Okuyama T., Shimizu A.;  
RT "Complete amino acid sequence of the Fc region of a human delta  
RL chain";  
RN Proc. Natl. Acad. Sci. U.S.A. 78:785-789(1981).  
RN [5]  
RN CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=81133632; PubMed=7008791;
```

```
RA Takayasu T., Takahashi N., Shinoda T.;  
RT "Amino acid sequence and location of the three glycopeptides in the  
RL Fc region of human immunoglobulin D";  
RL Biochem. Biophys. Res. Commun. 97:635-641(1980).  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
DR PIR; A02175; DHRU.  
DR Genew; HGNC:5480; IGHD.  
DR MIM; 147170; .  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003597; Ig-cl.  
DR InterPro; IPR003006; Ig_MHC.  
DR Pfam; PF00047; Ig; 3.  
DR SMART; SM00407; Igc1; 3.  
DR PROSITE; PS50835; IG-LIKE; 3.  
DR PROSITE; PS00290; IG_MHC; 2.  
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.  
FT NON TER 1  
FT DOMAIN 1 6 98 IG-LIKE 1.  
FT DOMAIN 175 263 IG-LIKE 2.  
FT DOMAIN 267 373 IG-LIKE 3.  
FT DISULFID 15 13 INTERCHAIN (WITH LIGHT CHAIN).  
FT DISULFID 28 84 INTERCHAIN (WITH HEAVY CHAIN).  
FT DISULFID 161 161  
FT CARBOHYD 109 109 O-LINKED.  
FT CARBOHYD 110 110 O-LINKED.  
FT CARBOHYD 113 113 O-LINKED.  
FT CARBOHYD 126 126 O-LINKED.  
FT CARBOHYD 127 127 O-LINKED.  
FT CARBOHYD 131 131 O-LINKED.  
FT CARBOHYD 132 132 O-LINKED.  
FT CARBOHYD 225 225 O-LINKED (POTENTIAL).  
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .).  
FT CARBOHYD 367 367 N-LINKED (GLCNAC. .).  
SQ SEQUENCE 383 AA; 42125 MW; DAIDE73519C76CID CRC64;  
  
Query Match 6.3%; Score 151.5; DB 1; Length 383;  
Best Local Similarity 23.5%; Pred. No. 0.0035;  
Matches 58; Conservative 32; Mismatches 88; Indels 69; Gaps 9;  
  
Qy 232 PKDTLMISRPETVCVVVDVSHEDPEVKFNWYDGEVFNHNAKTKPREQYNSTRYVSVL 291  
Db 18 PKD-----NSPVLACLI-TGYHPTSVTVTWY-GRSQSPQRIFPEIQRDSYMTSOL 70  
  
Qy 292 TVLHQDLNKEVKKYKSN-----KALPAP 316  
Db 71 STPLQWRQG-EYKCVQHTASKSKKEIFRWPSPYKAQASSVPYPAQAGSLAKATTAP 129  
  
Qy 317 I-----EKTISKAKGQPREPQ-----VYTLPPSDEL-TKNQVSLT 351  
Db 130 ATRNTRGGEFEKKEKEKEGEERETKTEPCSHNQPLGVLLTPAVQDLMLDKATFT 189  
  
Qy 352 CLVKGFPYSDIAVWESNGQ-PENNYKTPPVLDSDGFFLYSKLTVDKSRMOQGNVFC 410  
Db 190 CFVVGSDLKDAHLTWEVAGVPTGSGVEGLLEBHSNQSOGHSRLTLPRLMNAGTSVTC 249  
  
Qy 411 SYVHEAL 417  
Db 250 TLNHPSL 256  
  
RESULT 83  
KACA_RAT STANDARD; PRT; 106 AA.  
ID KACA_RAT  
AC P01836;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig kappa chain C region, A allele.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=DA;
RX MEDLINE=82082587; PubMed=6273908;
RA Sheppard H.W., Gutman G.A.;
RT "Allelic forms of rat kappa chain genes: evidence for strong
selection at the level of nucleotide sequence."
Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02118; KIRTA.
DR HSSP; P01842; 2MCG.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_1like; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
SQ SEQUENCE 106 AA; 11732 MW; B7E120D970DD66 CRC64;

Query Match 6.2%; Score 150; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.00088;
Matches 29; Conservative 23; Mismatches 28; Indels 12; Gaps 2;

QY 331 PVTYLPSPRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSE----- 385
DB 5 PVTYLPSPRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSE----- 385
DB 5 PVTYLPSPRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSE----- 385
QY 386 --DGSFELYSKLTVDKSRWQGNVFSGVMAE 415
DB 60 SKDSTYSMSSTSLTKVEYERHNLTYCEVYHK 91

RESULT 84
HB2C_PIG STANDARD; PRT; 261 AA.
AC P15982;
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SLA class II histocompatibility antigen, DQ haplotype C beta chain
DE precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sue.
OX NCBI_TaxId=9823;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=90361905; PubMed=2391424;
RA Gustafson K., Leguern C., Hirsch F., Germana S., Pratt K.,
RA Sachs D.H.;
RT "Class II genes of miniature swine. IV. Characterization and
expression of two allelic class II DQB cDNA clones."
J. Immunol. 145:1946-1951(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M31497; AAA31084.1; -
CC EMBL; M32117; AAA53110.1; -
CC PIR; A60404; A60404.
CC HSSP; P13760; 2SEB.

DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig_1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_1like; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 31
FT CHAIN 32 261 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
DO HAPLOTYPE C BETA CHAIN.
FT DOMAIN 127 220 EXTRACELLULAR BETA-2.
FT DOMAIN 221 230 CONNECTING PEPTIDE.
FT TRANSMEM 231 251 CYTOPLASMIC TAIL.
FT DOMAIN 252 261 BY SIMILARITY.
FT DISULFID 47 111 BY SIMILARITY.
FT DISULFID 149 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 51 51
SQ SEQUENCE 261 AA; 29550 MW; AA9581F2A3B1969D CRC64;

Query Match 6.2%; Score 149.5; DB 1; Length 261;
Best Local Similarity 23.2%; Pred. No. 0.0029;
Matches 56; Conservative 40; Mismatches 106; Indels 39; Gaps 8;

QY 180 GTWCTVLQNKQKVEFKIDIVCPAPRPSCKDHTCPBLGGPSVLPFPKPKDTLMIS 239
DB 12 GLMTAAL-----TWLVLVLAGPVAEGRD-----SPQDFVQFK-GEQFYFN 51
QY 240 RPEPTCVVDVSHDEPKRMVYDVEVNAKTKPEEQYNSTYRVSVLTVTHQWL 299
DB 52 GQRRRGVARYIYNQEHLRFDSDVEFRAYTPLEGRPADSNQKDLQWRA-EVDR 110
QY 300 NGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYP 359
DB 111 CKHNYQIE-----EGTTLQRRVP---TYTIPSKAEALNHNHNLVCAVTDYRP 156
QY 360 SDIAVEWESNGQPN-NYKTPPVLDSDGSFELYSKLTVDKSRWQGNVFSGVMAEALH 418
DB 157 SQVKVQWRNGQEBTAGVSTPLIRNGD---WTYQVLVLEMINLQRGDVYTCRVERHSLQ 213
QY 419 N 419
DB 214 N 214

RESULT 85
K1LO_RAT STANDARD; PRT; 348 AA.
AC Q920J8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kilon protein precursor (Kindred of IGLON).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-62.
RX MEDLINE=99175207; PubMed=10075727;
RA Funatsu N., Miyata S., Kumanooh H., Shigeta M., Hamada K., Endo Y.,
RA Sokawa Y., Maekawa S.;
RT "Characterization of a novel rat brain glycosylphosphatidylinositol-
anchored protein (Kilon), a member of the IGLON cell adhesion molecule
family."
J. Biol. Chem. 274:8224-8230(1999).
CC -1- FUNCTION: CELL-ADHESION (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: Highly expressed in brain.

```

CC -1- PTM: Glycosylated.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IGLON
CC family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB017139; BAA75649.1; -.
CC InterPro: IPR007110; IG_1-like.
CC InterPro: IPR003598; IG_C2.
CC Pfam: PR00047; Ig_3.
CC SMART, SM00408; IgC2; 2.
CC PROSITE: PS50835; IG_LIKE; 3.
CC Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
CC Repeat; Signal.
CC SIGNAL 1 31
CC CHAIN 32 ? KILON PROTEIN.
CC PROPE 348 REMOVED IN MATURE FORM (POTENTIAL).
CC DOMAIN 32 128 IG-LIKE C2-TYPE 1.
CC DOMAIN 133 215 IG-LIKE C2-TYPE 2.
CC DOMAIN 219 307 IG-LIKE C2-TYPE 3.
CC DISULFID 54 112 POTENTIAL.
CC DISULFID 154 197 POTENTIAL.
CC DISULFID 239 291 POTENTIAL.
CC CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 348 AA; 37858 MW; 37890DICTD24ACAB CRC64;

```

Query Match 6.2%; Score 149; DB 1; Length 348;
 Best Local Similarity 22.4%; Pred. No. 0.0045;
 Matches 72; Conservative 42; Mismatches 119; Indels 88; Gaps 14;

```

OY 20 PAATGKNVVLGKKGDTVELTCTASQKSIQFHWNISQIKILNQSGFLTKGP-----SK 75
DB 35 PMAADVNNMLV--RKGDITAVLRCLYEDGAS--KGAMLNRSIIIPAG--GDKMSVDPRVSI 89
OY 76 LNDRADSRSLMDQGNFLIILNKLIEDSDTYICGEVDQKEVQLLVFGLA----- 127
DB 90 LNKR-----DYSLIQIQNVVTDGPTVCSVQTOHTPRTMQVH-LTVQVPPKIID 137
OY 128 -NSDTHLQGSILTLTLESPGSPSPVOCR--SPRGKNIQGGKTLVSQLELDQSGTWTC 184
DB 138 ISNDMTINEGNTVITLTCLATGKPRPAISMRHISPAKFPENGQYDITGIRDOAGEIEC 197
OY 185 TV-----LQNKQKEFKIDIVCPAPEKSCDKHTTCPELLGGPSVFLPPKPKDTLMIS 239
DB 198 SAENDVSFPDYKKVAVVNFAP-----TIQEKISG-----TVPG 232
OY 240 RTEPTVTCVVVVDVSHDEPKFKWY-----VQGEVNHAKTTPREEQYNSTRVIVSV 290
DB 233 RSGLLRCGAGV---PPAPEWYGEKRLFENGQGIITQNFSTR-----SI 275
OY 291 LTVLHQDMLNGEKYKCKVSNK 311
DB 276 LVTNVTOEHFGNTYCVANAK 296

```

RESULT 86
 KAC_HUMAN STANDARD: PRT; 106 AA.
 AC P01834;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)

```

DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG kappa chain C region.
GN IGC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE (MELOMA PROTEIN EU).
RX MEDLINE=71064023; PubMed=5489770;
RA Gottlieb P.D., Cunningham B.A., Rutishauser U., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VI. Amino
RT acid sequence of the light chain."
RL Biochemistry 9:3155-3161(1970).
[2]
RP DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Gall W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RT intrachain disulfide bonds."
RL Biochemistry 9:3188-3196(1970).
[3]
RP SEQUENCE (BENCE-JONES PROTEIN TI).
RX MEDLINE=72188439; PubMed=5027703;
RA Suter L., Barnikol H.U., Watanabe S., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT immunoglobulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein
RT Ti). IV. The complete amino acid sequence and its significance for
RT the mechanism of antibody production."
RL Hoppe-Seyler's Z. Physiol. Chem. 353:189-208(1972).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=8104304; PubMed=6775818;
RA Hietter P.A., Max E.E., Seidman J.G., Maizel J.V. Jr., Leder P.;
RT "Cloned human and mouse kappa immunoglobulin constant and J region
RT genes conserve homology in functional segments."
RL Cell 22:197-207(1980).
[5]
RP SEQUENCE (BENCE-JONES PROTEIN ROY).
RX Hilschmann N., Barnikol H.U., Heas M., Langer B., Ponstingl H.,
RA Steimetz-Kayne M., Suter L., Watanabe S.;
RL (In) Franek F., Shugar D. (eds.);
RL Gamma globulins: structure and function, pp.57-74, Academic Press,
RL New York (1969).
[6]
RP SEQUENCE (BENCE-JONES PROTEIN CUM).
RX MEDLINE=68242259; PubMed=5586923;
RA Hilschmann N.;
RT "The complete amino acid sequence of Bence Jones protein Cum (kappa-
RT type)."
RL Hoppe-Seyler's Z. Physiol. Chem. 348:1718-1722(1967).
[7]
RP SEQUENCE (BENCE-JONES PROTEIN AG).
RX MEDLINE=69234734; PubMed=4893682;
RA Tltan K., Shinoda T., Putnam F.W.;
RT "The amino acid sequence of a kappa type Bence-Jones protein. 3. The
RT complete sequence and the location of the disulfide bridges."
RL J. Biol. Chem. 244:3550-3560(1969).
[8]
RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN OU).
RX MEDLINE=70201507; PubMed=5447531;
RA Kohler H., Shimizu A., Paul C., Putnam F.W.;
RT "Macroglobulin structure: variable sequence of light and heavy
RT chains."
RL Science 169:56-59(1970).
[9]
RP SEQUENCE OF 1-33; 38-41 AND 62-80.
RC TISSUE=Abdominal adipose tissue;
RX MEDLINE=98249779; PubMed=9588180;
RA Olsen K.E., Sletten K., Westermark P.;
RT "Extended analysis of AL-amyloid protein from abdominal wall
RT subcutaneous fat biopsy: kappa IV immunoglobulin light chain."
RL Biochem. Biophys. Res. Commun. 245:713-716(1998).

```

CC -1- MISCELLANEOUS: The EU sequence has the INV (3) allotypic marker,
CC 45-Mar and 83-Jul. The ROY sequence has the INV (1,2) allotypic
CC marker. 45-Mar and 83-Jul.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J00241; AAA58989.1; -
DR EMBL: V00557; CAA23823.1; -
DR PIR: B90562; K3HU.
DR PDB: 1D5B; 09-FEB-00.
DR PDB: 1D5I; 09-FEB-00.
DR PDB: 1D6V; 04-OCT-00.
DR PDB: 1HEZ; 18-JUL-03.
DR PDB: 1HKL; 12-MAR-97.
DR PDB: 117Z; 08-AUG-01.
DR PDB: 1M1M; 15-MAY-97.
DR GeneW: HGNC:5716; IGKC.
DR MIM: 147200; -
DR GO: 0003823; P:antigen binding; NAS.
DR GO: 0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-1like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00407; IGc1; 1.
DR PROSITE: PS00835; IG_Like; 1.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON TER 1 102 IG-Like.
FT DOMAIN 5 102
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT VARIANT 83 83 V -> L (IN INV(1,2) MARKER).
FT /FTID=VAR_003897.
FT CONFLICT 14 14 D -> N (IN REF. 7 AND 8).
FT CONFLICT 57 57 E -> Q (IN REF. 5 AND 6).
SQ SEQUENCE 106 AA; 11609 MW; 51984D1FDD372CE8 CRC64;
Query Match 6.1%; Score 148; DB 1; Length 106;
Best Local Similarity 29.5%; Pred. No. 0.0012;
Matches 28; Conservative 24; Mismatches 41; Indels 2; Gaps 1;
QY 331 PQVYTLPRSRDELTKNQVSLTCLVKGFYPSDIAVEWESNG--QPENNYKTTTPPVLDSDGS 388
DB 5 PSVFLFPSPDEQLKSGTASVCLLNPNFYREKVKQKVNALQSGNSQSEVTEDQSKST 64
QY 389 FFLYSKLTVDKSRMOQGNFSCSVMEALAHNYTQ 423
DB 65 YELSLTLLTSKADYEKHKYAAACEVTHGQSLSPVK 99
RESULT 87
NFAS CHICK STANDARD; PRT; 1369 AA.
AC 042414; 090924;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurofascin precursor.
GN NFASC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 26-46; 637-641; 717-730;
RP 758-781 AND 801-815, CLEAVAGE AT ARG-636, AND GLYCOSYLATION.
RC TISSUE=Brain;
RX MEDLINE=92317154; PubMed=1377696;
RA Volkmann H., Hassel B., Wolf J.M., Frank R., Rathjen F.G.;
RT "Structure of the axonal surface recognition molecule neurofascin and
RT its relationship to a neural subgroup of the immunoglobulin
RL J. Cell Biol. 118:149-161(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Liver;
RX MEDLINE=98019255; PubMed=9353344;
RA Hassel B., Rathjen F.G., Volkmann H.;
RT "Organization of the neurofascin gene and analysis of developmentally
RT regulated alternative splicing";
RL J. Biol. Chem. 272:28742-28749(1997).
CC -1- FUNCTION: Cell adhesion, ankyrin-binding protein which may be
CC involved in neurite extension, axonal guidance, synaptogenesis,
CC myelination and neuron-glia cell interactions (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing: Named isoforms=2;
CC Comment=A number of isoforms are produced;
CC Name=1;
CC IsoId=042414-1; Sequence=Displayed;
CC Name=2;
CC IsoId=042414-2; Sequence=VSP_008935, VSP_008936;
CC -1- DEVELOPMENTAL STAGE: There is one major 'early' isoform and
CC multiple 'late' isoforms. Around 50 isoforms are found at
CC different developmental stages.
CC -1- PTM: N-glycosylated and O-glycosylated.
CC -1- PTM: May be proteolytically cleaved at ARG-636.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC L1/neurofascin/NgCAM family.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65224; CAA46330.1; -
DR EMBL: Y14341; CAA74726.1; -
DR EMBL: Y14342; CAA74726.1; JOINED.
DR EMBL: Y14343; CAA74726.1; JOINED.
DR EMBL: Y14344; CAA74726.1; JOINED.
DR EMBL: Y14345; CAA74726.1; JOINED.
DR EMBL: Y14346; CAA74726.1; JOINED.
DR EMBL: Y14347; CAA74726.1; JOINED.
DR EMBL: Y14348; CAA74726.1; JOINED.
DR EMBL: Y14349; CAA74726.1; JOINED.
DR EMBL: Y14350; CAA74726.1; JOINED.
DR EMBL: Y14351; CAA74726.1; JOINED.
DR EMBL: Y14352; CAA74726.1; JOINED.
DR EMBL: Y14353; CAA74726.1; JOINED.
DR EMBL: Y14354; CAA74726.1; JOINED.
DR PIR: S26180; S26180.
DR HSSP: P20241; 1CFB.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR007110; IG-1like.
DR InterPro: IPR003598; IG_c2.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Ig; 6.
DR SMART: SM00060; FN3; 5.
DR SMART: SM00408; IGc2; 3.
DR PROSITE: PS00835; IG_Like; 6.
KW Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain;

KW Glycoprotein; Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1369
 FT DOMAIN 26 1235
 FT TRANSMEM 1236 1286
 FT DOMAIN 1257 1369
 FT DOMAIN 42 138
 FT DOMAIN 144 231
 FT DOMAIN 252 350
 FT DOMAIN 355 442
 FT DOMAIN 448 535
 FT DOMAIN 539 626
 FT DOMAIN 643 735
 FT DOMAIN 742 835
 FT DOMAIN 840 942
 FT DOMAIN 943 1052
 FT DOMAIN 1134 1207
 FT DOMAIN 1052 1126
 FT SITE 636 637
 FT DISULFID 64 119
 FT DISULFID 163 214
 FT DISULFID 286 334
 FT DISULFID 376 426
 FT DISULFID 470 519
 FT DISULFID 561 610
 FT CARBOHYD 241 241
 FT CARBOHYD 247 247
 FT CARBOHYD 323 323
 FT CARBOHYD 427 427
 FT CARBOHYD 464 464
 FT CARBOHYD 501 501
 FT CARBOHYD 692 692
 FT CARBOHYD 767 767
 FT CARBOHYD 793 793
 FT CARBOHYD 853 853
 FT CARBOHYD 994 994
 FT CARBOHYD 1009 1009
 FT CARBOHYD 1133 1133
 FT CARBOHYD 1150 1150
 FT CARBOHYD 1156 1156
 FT CARBOHYD 1171 1171
 FT VARSPLIC 967 993
 FT VARSPLIC 1132 1222
 FT CONFLICT 1105 1105
 FT SEQUENCE 1369 AA; 152954 MM; 835f27D086B2BF6 CRC64;
 SQ
 Query Match 6.1%; Score 148; DB 1; Length 1369;
 Best Local Similarity 22.3%; Pred. No. 0.029;
 Matches 107; Conservative 63; Mismatches 182; Indels 128; Gaps 31;
 4 GVPF---RHLLLVQLAL-----LP---AAQGNKVVLGKGGVVELTCTASQKKS 48
 12 GIAFALCLHHLISALEVPLDSNIGSELPPPTTQSVYDVDRDNIIECECKGNPV 71
 49 IOFHKNNSNQIKILNOSGF--LTGSPKSLNDRADSRSLMD-----QGNFPLII 96
 72 PTFSTWTRN-----GKFENVAADP-KVSMRRKRSGLVLVDFFHGGSPRDYEGEYQCPA 121
 97 KN-----LKIEDSDTYICEVEDOKEEVQLLVGLTANSDTHLLQGOSLTLLLES 145
 122 RNDYGTALSKIHLYGVSRSPPLW-----PKERKVDYI-----EVDGAPLISLQGNP 165
 146 PPGSSPSV-----QCRSP--RKNIGGKGT--LSISQLELDQSGT-WICT-----VLQ 188
 166 PPGLPPIVIFMWSMSEPIHODKRVSOQGNBDLYSNWMLDPAQDYSCNARFHTHTIQ 225
 189 NOKKVEFKIDIVPCAPBEPKSCDKTHTCELLGG-----PSVFLRPPPKDTLMISRTPE 243
 226 QKNPTTLK---VKTKKPHNETISLRNHTDYSARGVTEPTTPS-FMTPYGTSSSQWLVLRVD 281
 244 --VTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREQYNSTYRVVSVLTVLHQDWLNG 301

Db 282 LLEECIASGV-----PADDIMWYKKGELPAKTK--LENFNKALRISNVSEE-----DS 329
 Qy 302 KEYCKCKSNKALPAPIKITS-KAKGP---REQVYTLPPSPDELTKNOVSLCYKGF 357
 Db 330 GEYFCLASNKK--GSINHTISVRKAPWLDPEQNIILAPGEDG-----RLVCRRANG- 380
 Qy 358 YPSDIKAVESNCGPENNYKTPPEVLDS--GSFFLSKLTVDKSRWQGNVFSQVMHE 415
 Db 381 NPKR-SIGMLVNGEP---IGSPNPSPREVAAGDITVRDQIGSSA-----VYGCNANSNE 431
 RESULT 88
 ID LAMP HUMAN STANDARD; PRT; 338 AA.
 AC 01343;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Limbic system-associated membrane protein precursor (LSAMP).
 GN LSAMP OR LAMP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96235133; PubMed=8666243.
 RA Pimenta A.F., Fischer I., Levitt P.;
 RT "cDNA cloning and structural analysis of the human limbic-system-
 associated membrane protein (LAMP).";
 RL Gene 170:189-195 (1996).
 CC -1- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
 CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
 CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
 CC OF THE HIPPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: Expressed on limbic neurons and fiber tracts
 CC as well as in single layers of the superior colliculus, spinal
 CC chord and cerebellum.
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
 CC family.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL, U41901, AAC50569.1, -.
 CC PIR, J04776; J04776.
 CC Genew, HGNC:6705; LSAMP.
 CC MIM, 603241; -.
 CC GO: GO:0007399; P:neurogenesis; TAS.
 CC InterPro: IPR007110; Ig-like.
 CC InterPro: IPR003596; Ig_c2.
 CC Pfam, PF00047; Ig; 3.
 CC SMART, SM00408; IgC2; 2.
 CC PROSITE, PS50835; IG LIKE; 3.
 CC Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
 KW Repeat; Signal; Lipoprotein.
 FT SIGNAL 1 28
 FT CHAIN 29 315
 FT FT
 FT PROPEP 316 338
 FT DOMAIN 29 122
 FT DOMAIN 132 214
 FT DOMAIN 219 304
 FT DISULFID 53 111
 FT DISULFID 153 197
 FT POTENTIAL.
 FT LIMBIC SYSTEM-ASSOCIATED MEMBRANE
 FT PROTEIN.
 FT REMOVED IN MATURE FORM (POTENTIAL).
 FT IG-LIKE C2-TYPE 1.
 FT IG-LIKE C2-TYPE 2.
 FT IG-LIKE C2-TYPE 3.
 FT POTENTIAL.

FT	DISTLPTD	239	290	POTENTIAL.
FT	CARBOHYD	40	40	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	66	66	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	136	136	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	148	148	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	279	279	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	287	287	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	300	300	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	315	315	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	Lipid	315	315	GPI-anchor amidated asparagine (Potential).
SQ	SEQUENCE	338 AA:	37308 MW;	03455F286DF592F CRC64;
	Query Match	6.1%;	Score 147;	DB 1; Length 338;
	Best Local Similarity	22.7%;	Pred. No. 0.0059;	
	Matches	75; Conservative	49; Mismatches	129; Indels 78; Gaps 16;
Oy	10 LLVLVQLLLPLRA-----TGQKRVVLGKKGDVELTCTASQCKSIOFHMKNSNOIKI	61		
Dd	14 LVLRRLCLLPGLPLRVSVDENRGDNTTVRCGDPAILRCVLEDNKS-KVALNRSGIIF	72		
Oy	62 LGNGSFPTKGPSKINDRA--DSRSLLMDGNFPILIKNLKIEDSDTYICEVEDKE---	116		
Dd	73 AGHD-----KMSLDPRVELEKHSLS---EYSLRIQKVDDYGSGSTCSVOJOHEKT	121		
Oy	117 -EVOLIVEG---LTNSTDTHLLOQSJTLTLESPPGSSPSVOCR--SPRGKNITOGKT-	168		
Dd	122 SQVYLIVGVPRKISMISSDDVTYNESANTLVCMANGREPEVIYWHHLPTGRREFGEDEY	181		
Oy	169 LSVGLELQDSTWTCTVLQ----NQKKVEFKIDIVCPAPRPSCSKTKTCPELLGP	223		
Dd	182 LEILDITEQGSGECKAEKANSADVQVKVTAYVP-TITESKSNAT-----	230		
Oy	224 SVFLPFRPKDTLMISRTPETVCVVVDVSHEDPEKFMMYYDGVEVNHA---KTRPREQ	280		
Dd	231 -----TGROASLKCEASAV---PADFEWRYDRDTRINSANGLKISTEQ	272		
Oy	281 YNSTYRVVSVLTVLHQDWLNGKEYKCKVSNK	311		
Dd	273 SSLT-----VTNVTEEHYGN---YTCVAANK	295		
	RESULT 89			
	NPAS_MOUSE			
ID	_NPAS_MOUSE	STANDARD;	PRT;	1240 AA.
AC	OS10U3:			
DT	15-MAR-2004 (Rel. 43, Created)			
DT	15-MAR-2004 (Rel. 43, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
Dd	Neurofascin precursor.			
GN	NPASC.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Swiss Webster; TISSUE=Brain;			
RA	Dixke P., Montag-Sallaz W., Montag D.;			
RT	"Expression patterns of Ll-family cell recognition molecules Ll, CHL1,			
RT	NrcAM, and neurofascin in the mouse brain.";			
RL	Submitted (FEB-2003) to the EMBL/Genbank/DDBJ databases.			
CC	- FUNCTION: Cell adhesion, ankyrin-binding protein which may be			
CC	involved in neurite extension, axonal guidance, synaptogenesis,			
CC	myelination and neuron-glia cell interactions (By similarity).			
CC	- SUBUNIT: Probable constituent of a neurofascin/NRCAM/ankyrin G			
CC	complex. Associates with the sodium channel beta-1 (SCN1B) and			
CC	beta-3 (SCN3B) subunits (By similarity).			
CC	- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	- SIMILARITY: Belongs to the immunoglobulin superfamily.			
CC	Ll/neurofascin/NRCAM family.			
-1-	SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.			
CC	-1- SIMILARITY: Contains 4 fibronectin type III domains.			

	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on ways CC can use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi.ac.uk/termsandconditions or send an email to license@ebi.ac.uk).
	DR	EMBL; AJ543322; CAD65849.1; -
	DR	MED; MG1:2442229; P430023G6R.RK.
	DR	InterPro; IPR003961; FN_III-like.
	DR	InterPro; IPR008957; FN_III-like.
	DR	InterPro; IPR003599; IG.
	DR	InterPro; IPR007110; IG-like.
	DR	InterPro; IPR003598; IG_c2.
	DR	Pfam; PF00041; fn3; 4.
	DR	Pfam; PF00047; Ig; 6.
	DR	SMART; SM00060; FN3; 4.
	DR	SMART; SM00409; IG; 6.
	DR	SMART; SM00408; IGc2; 6.
	DR	PROSITE; PS50835; IG_LIKE; 6.
	KW	Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein.
	KM	GLYCOPROTEIN.
	FT	SIGNAL 1 24 POTENTIAL.
	FT	CHAIN 25 1240 NEUROFASCIN.
	FT	DOMAIN 25 1110 EXTRACELLULAR (POTENTIAL).
	FT	TRANSMEM 1111 1131 POTENTIAL.
	FT	DOMAIN 1132 1240 CYTOPLASMIC (POTENTIAL).
	FT	DOMAIN 41 137 IG-LIKE C2-TYPE 1.
	FT	DOMAIN 143 230 IG-LIKE C2-TYPE 2.
	FT	DOMAIN 244 332 IG-LIKE C2-TYPE 3.
	FT	DOMAIN 337 424 IG-LIKE C2-TYPE 4.
	FT	DOMAIN 430 517 IG-LIKE C2-TYPE 5.
	FT	DOMAIN 521 603 IG-LIKE C2-TYPE 6.
	FT	DOMAIN 628 720 FIBONECTIN TYPE-III 1.
	FT	DOMAIN 727 820 FIBONECTIN TYPE-III 2.
	FT	DOMAIN 825 924 FIBONECTIN TYPE-III 3.
	FT	DOMAIN 1008 1094 FIBONECTIN TYPE-III 4.
	FT	DOMAIN 924 1006 THR-RICH.
	FT	DISULFID 63 118 POTENTIAL.
	FT	DISULFID 162 213 POTENTIAL.
	FT	DISULFID 268 316 POTENTIAL.
	FT	DISULFID 358 408 POTENTIAL.
	FT	DISULFID 452 501 POTENTIAL.
	FT	DISULFID 543 592 POTENTIAL.
	FT	CARBOHYD 305 305 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD 409 409 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD 446 446 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD 483 483 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD 752 752 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD 778 778 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD 866 866 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD 881 881 N-LINKED (GLCNAC . .) (POTENTIAL).
	SQ	SEQUENCE 1240 AA; 137975 MW; 6DBE935B502EB95 CRC64;
	Query March	6.1%; Score 147; DB 1; Length 1240;
	Best Local Similarity	21.4%; Pred. No. 0.03; Mismatches 183; Indels 118; Gaps 26
	Matches 100; Conservative	66; Mismatches 183; Indels 118; Gaps 26
Dy	6	PERHLVLQLALPA-----TGCKNVVIGKKGDVELTCTTSQ 45
Dy	8	PWVAIALILFLSLSGAIEIPMDPSIGNELTPPTITQSVDHIVDPBDIILIECEAKG 67
Dy	46	KKSIOFTW-KNSNOIKILNGCSFLTK--GPSKLNDRAISRSLMDGNFPLITKN---- 98
Dy	68	NPAFSFWTRSREFNIADPRVAMRRRSRGTLVIDFRSGGRPEY-EGEYOCFANKRKGT 126
Dy	99	-----LKIEDSTPIYCEVEDQKEVEQVLVFGLTANSDFHLLDGOSLTLLBSPPG-SS 150
Dy	127	ALSNRIRIQVSKSPW-----PRENDPVV-----VQGAPYLTCNPPGPLPS 170
Dy	151	PSVCCRSPRGKNIGGKTLS-----VSOLELDQSGT-WTCVLAQN-----OKVAFK 196

```

Db      171. PFIFWSSSSMERITODKRVSSQGNGLDYFSNWMLQDMQTDYSCNARFHFHTHTIOQKNEFT 230
      197 IDIVCPAPKPEPKCKHTCTCPCLLGGSPVFLPFPKXTOLMISRPPE--VTCVVVDVSHE 254
      221 LKVLTTTGVAEKTE-----PS-FMTPGCTSSQVWLRMDLLLECISGV--- 273
      255 DPEVFNMYVDGVEVHNAKTKPREQVNSTYRVSVLTVLVQDMWLNKGEYKCKVSNKALP 314
      274 -PTPIAIYKKKGGLPSNKAK--FENFKALRITVSE-----DSGEYICLASNMK-- 322
      315 APIETKTS-KAKGP--REPOVYTLPPSRDELTKNOVSLTCLVKGYFSPSDIAVWESNG 370
      323 GSIRHTISVRYKAAFYMLDEPKNLLIAPGSDG-----RLVCRANG-NPKP-TVQMWVNG 374
      375 EP--LQSAFPPVNPNEVAGDTTIFRDTOISSRA-----VYQCNISNE 413

```

RESULT 90

LAC1_RAT STANDARD; PRT; 104 AA.

```

AC P20766;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-1 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87305594; PubMed=3114047;
RA Steen M.L., Hellman L., Petersen U.;
RT "The immunoglobulin lambda locus in rat consists of two C lambda
RT genes and a single V lambda gene."
RL Gene 55:75-84(1987).

```

-1- SIMILARITY: Contains 1 immunoglobulin-like domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; M22520; AAA41419.1; ALT_INIT.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGc1.1.
DR PROSITE; PS50835; IG LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT DOMAIN 1
FT DISULFID 27
FT DISULFID 103
SQ SEQUENCE 104 AA; 11565 MW; CBF71811F4BC878A CRC64;

```

Query Match 6.0%; Score 146; DB 1; Length 104;
 Best Local Similarity 33.0%; Pred. No. 0.0016;
 Matches 34; Conservative 19; Mismatches 44; Indels 6; Gaps 4;

```

Oy      327 QPR-EPQVYTLPPSRDELTKNOVSLTCLVKGYFSPSDIAVWESNGQP-ENNYKTPPYLD 384
      1 QPKATPSVTLPPSSSEELTKATILVCWVTDFYGVWTVWKAADSTPIQGVETIQP-FK 59

```

```

Oy      385 SDGSFPLYSKLTVDKSRMOQGNVFSQVMEALHNHYTOKSLS 427
      60 QNNKYMATSYLLLTAKAMETHSNVSCQVTHE---ENTVERSLIS 99

```

RESULT 91

CAML_RAT STANDARD; PRT; 1259 AA.

```

AC O05695;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule L1 precursor (N-CAM L1).
GN L1CAM OR CAML1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=91372414; PubMed=1894011;
RA Miura M., Kobayashi M., Asou H., Uyemura K.;
RT "Molecular cloning of cDNA encoding the rat neural cell adhesion
RT molecule L1. Two L1 isoforms in the cytoplasmic region are produced
RT by differential splicing."
RL FEBS Lett. 289:91-95(1991).

```

-1- FUNCTION: Cell adhesion molecule with an important role in the development of the nervous system. Involved in neuron-neuron adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds to axonin on neurons.

-1- SUBCELLULAR LOCATION: Type I membrane protein.

-1- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=1;

isoId=Q05695-1; Sequence=displayed;

Name=2; Synonyms=L1cs;

isoId=Q05695-2; Sequence=VSP_002592;

-1- TISSUE SPECIFICITY: Isoform 2 is predominantly found in the brain, while isoform 1 is found in the peripheral nervous system.

-1- SIMILARITY: Belongs to the immunoglobulin superfamily.

-1- neurofascin/NgCAM family.

CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.

CC -1- SIMILARITY: Contains 5 fibronectin type III domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; X59149; CAA41860.1; -.
DR PIR; S36126; S36126.
DR HSSP; P20241; ICFB.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00041; Fn3; 4.
DR Pfam; PF00047; Ig; 6.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS50835; IG LIKE; 6.
KW Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;
KW Transmembrane; Repeat; Immunoglobulin domain; Signal;
KW Alternative splicing.
FT SIGNAL 1
FT CHAIN 20
FT DOMAIN 20
FT TRANSMEM 1123
FT DOMAIN 1146
FT DOMAIN 1146
FT DOMAIN 1259
FT DOMAIN 1259

```

BY SIMILARITY.
 NEURAL CELL ADHESION MOLECULE L1.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 IG-LIKE C2-TYPE 1.


```

RP SEQUENCE FROM N.A.
RC STRAIN=Louvain;
RX MEDLINE=82082587; PubMed=62733908;
RA Sheppard H.W., Gutman G.A.;
RT "Allelic forms of rat kappa chain genes: evidence for strong
RL selection at the level of nucleotide sequence.";
RN Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).
[2]
RP SEQUENCE (BENCE-JONES PROTEIN S211).
RX MEDLINE=75212238; PubMed=807630;
RA Statace V., Querinjean P.;
RT "The primary structure of a rat kappa Bence Jones protein:
RT phylogenetic relationships of V- and C-region genes.";
CC J. Immunol. 115:59-62(1975).
-1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A83901; K1RFB.
DR HSP, P01842; 2MCG.
DR InterPro; IPR007110; Ig-1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.1.
DR SMART; SM00407; Ig1.1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT CONFLICT 2 2 D -> N (IN REF. 2).
FT CONFLICT 30 30 N -> K (IN REF. 2).
FT CONFLICT 48 48 MISSING (IN REF. 2).
FT CONFLICT 79 79 E -> Q (IN REF. 2).
FT CONFLICT 87 87 E -> O (IN REF. 2).
FT CONFLICT 98 98 V -> W (IN REF. 2).
FT CONFLICT 100 100 S -> N (IN REF. 2).
SQ SEQUENCE 106 AA; 11601 MW; 4CFA7CA820D1CA36 CRC64;

Query Match 6.0%; Score 145; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.0019;
Matches 29; Conservative 22; Mismatches 29; Indels 12; Gaps 2;

QY 331 POYVTLPSRDELTKNQVSLTCLVKGYFSPYSDIAVESNGPENNYKTPPYLDS----- 385
DB 5 PTVSIFPSTEQLATGASVVCVMNNFYPRDISVXWKIDGERD-----GVLDSVTDD 59
QY 386 --DGSFPLYSKLTVDKSRMOCNVFSCVME 415
DB 60 SKDSTYSMSSTLSLTKADYESHNLTYCEVHK 91

RESULT 94
AMAL_DROME STANDARD; PRT; 333 AA.
ID AMAL_DROME STANDARD; PRT; 333 AA.
AC P15364; Q9V3A5;
DC 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Amalgam protein precursor.
GN AMA OR BG:DS00276.6 OR CG2198.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=89028670; PubMed=3141062;
RA Seeger M.A., Hafley L., Kaufman T.C.;
RT "Characterization of amalgam: a member of the immunoglobulin
RL superfamily from Drosophila.";
RL Cell 55:589-600(1988).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Celniker S.E., Pfeiffer B.D., Kafets J., Martin C.H., Mayeda C.A.,
RA Palazzolo M.J.;
RT "Complete sequence of the Antennapedia complex of Drosophila.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotiler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
RA Jalaal M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Moritz G., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yelt R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
-1- (potential).
-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; M23561; AAA28367.1; -.
CC EMBL; AE001572; AAD19797.1; -.

```

DR EMBL; AE003674; AAF54084.1; -
 DR EMBL; AY051911; AAK93335.1; -
 DR PIR; A31923; A31923.
 DR FlyBase; Fgn0000071; Ama.
 DR GO; GO:0005886; C:plasma membrane; IDA.
 DR InterPro; IPR007110; IG-1~~4~~C.
 DR InterPro; IPR003598; IG_C2.
 DR Pfam; PF00047; Ig_3.
 DR SMART; SM00408; IGC2; 2.
 DR PROSITE; PS50835; IG_LIKE; 3.
 KW Immunoglobulin domain; Glycoprotein; Membrane; GPI-anchor; signal;
 Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 ? AMALGAM PROTEIN
 FT PROSEP 2 333 REMOVED IN MATURE FORM (POTENTIAL).
 FT DOMAIN 25 128 IG-LIKE V-TYPE.
 FT DOMAIN 139 223 IG-LIKE C2-TYPE 1.
 FT DOMAIN 230 323 IG-LIKE C2-TYPE 2.
 FT DISULFID 46 117 PROBABLE.
 FT DISULFID 161 208 PROBABLE.
 FT DISULFID 251 307 PROBABLE.
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 83 83 O -> K (IN REF. 1).
 SQ SEQUENCE 333 AA; 36387 MW; F644753DB3DB25F1 CRC64;

Query Match 6.0%; Score 144.5; DB 1; Length 333;
 Best Local Similarity 19.9%; Pred. No. 0.0086;
 Matches 70; Conservative 57; Mismatches 127; Indels 97; Gaps 14;

QY 10 LLLVLQAL-----LPATQGNKVVLGKGDVVELTCTASQKKSIOFHW-----KNSNQ 58
 DB 10 LFFCLALSDSVLSAVVISQISDVVASVDSVEFNCVTEEVGQSVSAKPPSSSDTNS 69
 QY 59 IKI-----LGNO--GSFLTKGPKSLANDRSRLSDQGNFPLIKLIKLIKIEDSTYIC 109
 DB 70 VLSNRNLTSLPDQRNVNTVTEGPKT-----GSATYTRIQNIIEVSDMGPRYC 117
 QY 110 EV-----EDQKEEVQLI--VFGLTANSSTHLLQGGSLTLTLESPGSSPSV----- 153
 DB 118 QVLVATEKVKTKKLSIQIKTPPVIAENTPKSLVTEGQNLLETCHANGFPKPTISWAEH 177
 QY 154 QGRSPRGKNIQGGKTLVSQLELDSCGTCTVNLQ-----QKVEEKIDIVPER 203
 DB 178 NAWMPAGGLLAEPLTKRISVIRMDGGYVC-IAQEGQDPKRLIRVEVEFRPOIA--- 233
 QY 204 ABEPRKCDKHTCPPELLGSPVFLPPPKPKDTLMI SRTEPVTCVVVDVSHEDPEVKFMY 263
 DB 234 -----VGRPKIAMQVSHSLALCSQVGY-----PAPTVVMH 264
 QY 264 VDGVEVHNAKYKPREQVNSTYRVVSVL--TVLHODWINGKEYKCKVSNK 311
 DB 265 KNGVPLQSSRHHEVANTASSGTTTSVLRLDSVGEEDF---GDYCNATNK 312

RESULT 95
 VCAM_MOUSE STANDARD; PRT; 739 AA.
 ID VCAM_MOUSE
 AC P29533;
 DT 01-APR-1993 (rel. 25, Created)
 DT 01-APR-1993 (rel. 25, Last sequence update)
 DT 10-OCT-2003 (rel. 42, Last annotation update)
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
 OS VCAM1 OR VCAM-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=FVB; TISSUE=Lung;
 RX MEDLINE=92181437; PubMed=1371918;

RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wyek M.,
 RA Burkly L., Miyake K., Kincade P., Lobb R.;
 RA "Cloning of murine and rat vascular cell adhesion molecule-1";
 RL Biochem. Biophys. Res. Commun. 183:163-169(1992).
 RL (2)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RP TISSUE=Lymph node;
 RX MEDLINE=93346254; PubMed=7683304;
 RA Araki M., Araki K., Vassalli P.;
 RT "Cloning and sequencing of mouse VCAM-1 cDNA";
 RL Gene 126:261-264(1993).
 RL (3)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RP STRAIN=129; TISSUE=Embryo;
 RX MEDLINE=94117008; PubMed=7507076;
 RA Cybulsky M.I., Allan-McLamed M., Collins T.;
 RT "Structure of the murine VCAM1 gene";
 RL Genomics 18:387-391(1993).
 RL (4)
 RP SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).
 RP STRAIN=NIH Swiss, and 129/Sv;
 RA Kumar A.G., Dai Y.X., Kozak C.A., Mims M.P., Gotto A.M. Jr.,
 RA Ballantyne C.M.;
 RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
 RL (5)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RP STRAIN=FVB; TISSUE=Lung;
 RX MEDLINE=93232042; PubMed=7682556;
 RA Moy P., Lobb R., Tizard R., Olson D., Hession C.;
 RT "Cloning of an inflammation-specific phosphatidylinositol-linked
 RT form of murine vascular cell adhesion molecule-1";
 RL J. Biol. Chem. 268:8835-8841(1993).
 RL (6)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RP STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=95015899; PubMed=7523515;
 RA Kumar A.G., Dai X.Y., Kozak C.A., Mims M.P., Gotto A.M.,
 RA Ballantyne C.M.;
 RT "Murine VCAM-1. Molecular cloning, mapping, and analysis of a
 RT truncated form";
 RL J. Immunol. 153:4088-4098(1994).
 RL (7)
 RP SEQUENCE OF 311-345 FROM N.A. (ISOFORM 2).
 RP STRAIN=FVB/N; TISSUE=Kidney;
 RX MEDLINE=93317595; PubMed=7687058;
 RA Terry R.W., Kwee L., Levine J.F., Labow M.A.;
 RT "Cytokine induction of an alternatively spliced murine vascular cell
 RT adhesion molecule (VCAM) mRNA encoding a
 RT glycosylphosphatidylinositol-anchored VCAM protein";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5919-5923(1993).
 RL (8)
 RP SEQUENCE OF 1-21 FROM N.A.
 RP TISSUE=Endothelial cells;
 RA Korenaga R., Ando J., Teubol H., Kamiya A.;
 RL Submitted (Dec-1995) to the EMBL/Genbank/DBJ databases.
 RL (9)
 RP FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
 RP IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
 RP INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
 RP TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
 RP PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
 RP EMIGRATION TO SITES OF INFLAMMATION.
 RP SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG ISOFORM) OR
 RP ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (SHORT ISOFORM).
 RP ALTERNATIVE PRODUCTS:
 RP Name=1; Synonyms=Long;
 RP IsoId=P29533-1; Sequence=Displayed;
 RP Name=2; Synonyms=Short;
 RP IsoId=P29533-2; Sequence=VSP_002581, VSP_002582;
 RP TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
 RP well as on macrophage-like and dendritic cell types in both normal
 RP and inflamed tissue.
 RP -!- PNM: The GPI-anchor is located on position 319 of isoform 2.


```

CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC OF THE HYPOCAMPAL MOSSY FIBER PROTECTION.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: Expressed mostly by neurons comprising limbic-
CC associated cortical and subcortical regions that function in
CC cognition, emotion, memory, and learning.
CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT E15-16, AT STAGE E20 IT IS
CC DETECTED IN PRESUMPTIVE CORTEX. MEDIAL LIMBIC AREAS OF THE
CC THALAMUS AND HYPOTHALAMUS. IN THE ADULT, IT IS FOUND IN
CC HYPOTHALAMUS, PERIRHINAL CORTEX, AMYGDALA AND MEDIAL THALAMIC
CC REGION.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. Iglon
CC family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U1554; AAA6120.1; -.
CC InterPro: IPR007110; Ig_Like.
CC InterPro: IPR003598; Ig_C2.
CC Pfam: PF00047; Ig_3.
CC SMART: SM00408; IgC2; 2.
CC PROSITE: PS50835; IG_Like; 3.
CC Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
CC Repeat; Signal; Lapporetein.
CC SIGNAL 1 38
CC CHAIN 29 315
CC -----
CC PROPEP 316 338 LIMBIC SYSTEM-ASSOCIATED MEMBRANE
CC DOMAIN 29 122 REMOVED IN MATURE FORM (POTENTIAL).
CC DOMAIN 132 214 IG-LIKE C2-TYPE 1.
CC DOMAIN 219 304 IG-LIKE C2-TYPE 2.
CC DISULFID 53 111 POTENTIAL.
CC DISULFID 153 197 POTENTIAL.
CC DISULFID 239 290 POTENTIAL.
CC CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC LIPID 315 315 GPI-anchor amidated asparagine
CC (POTENTIAL).
CC SEQUENCE 338 AA; 37324 MW; 0B76AFD68A39B86 CRC64;
CC -----
CC Query Match 6.0%; Score 144; DB 1; Length 338;
CC Best Local Similarity 22.4%; Pred. No. 0.0095;
CC Matches 74; Conservative 50; Mismatches 129; Indels 78; Gaps 16;

```

```

QY 224 SVFLPPPKKDKTLMISRTPEVTCVVVDVSHEDPEVFNNYDVGVEHNA---KTKPREQ 280
DB 231 -----TGRQASLKCENSAV-----PADDFEYRDRDTKINSANGLEISTEQ 272
QY 281 YNSTYRVSVLTLVTHQDMNGEKYKCVSNK 311
DB 273 SSLT-----VTNVETEHYGN---YTCVANK 295
RESULT 97
NFAS RAT
ID NFAS RAT STANDARD; PRT; 1240 AA.
AC P97685; P97684; O91260;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurofascin precursor.
GN NFASC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC STRAIN=Wistar;
RA Tate S., Collinson J.M., Brophy P.J.;
RN Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 25-1240 FROM N.A. (ISOFORMS 1; 2 AND 3), AND SUBUNIT.
RC TISSUE=Brain;
RX MEDLINE=97103184; PubMed=8947556;
RA Davis J.Q., Lambert S., Bennett V.;
RT "Molecular composition of the node of Ranvier: identification of
RT ankyrin-binding cell adhesion molecules neurofascin (mucin+/third
RT FNIII domain-) and NrCAM at nodal axon segments.";
RN J. Cell Biol. 135:1355-1367(1996).
[3]
RP FUNCTION OF ISOFORM 2/3.
RX MEDLINE=98220650; PubMed=9562181;
RA Collinson J.M., Marshall D., Gillespie C.S., Brophy P.J.;
RT "Transient expression of neurofascin by oligodendrocytes at the onset
RT of myelogenesis: implications for mechanisms of axon-glia
RT interaction.";
RN J. Cell Biol. 135:1355-1367(1996).
[4]
RP FUNCTION OF ISOFORMS 1 AND 2/3. INDIRECT ASSOCIATION WITH CNTNAP1,
RP PHOSPHORYLATION, AND TISSUE SPECIFICITY.
RX MEDLINE=20391985; PubMed=10931875;
RA Tate S., Gunn-Moore F., Collinson J.M., Huang J., Lubetzki C.,
RA Pedraza L., Sherman D.L., Colman D.R., Brophy P.J.;
RT "An oligodendrocyte cell adhesion molecule at the site of assembly of
RT the paranodal axo-glia junction.";
RN J. Cell Biol. 150:657-666(2000).
[5]
RP INTERACTION WITH SODIUM CHANNEL BETA-1 AND BETA-3 SUBUNIT.
RX MEDLINE=21363577; PubMed=11470829;
RA Ratcliffe C.F., Westendorp R.E., Curtis R., Catterall W.A.;
RT "Sodium channel beta1 and beta3 subunits associate with neurofascin
RT through their extracellular immunoglobulin-like domain.";
RN J. Cell Biol. 154:427-434(2001).
-1- FUNCTION: Cell adhesion, ankyrin-binding protein which may be
CC involved in neurite extension, axonal guidance, eynaplogenesis,
CC myelination and neuron-glia cell interactions. Isoforms 2/3 may
CC be responsible for mediating and signaling axon-glia interaction
CC during the early stages of myelination.
CC -1- SUBUNIT: Probable constituent of a neurofascin/NrCAM/ankyrin G
CC complex. Associates with the sodium channel beta-1 (SCN1B) and
CC beta-3 (SCN1B) subunits. Associates to beta-1 subunit in
CC developing axons as early as postnatal day 5, during the period
CC that nodes of Ranvier are forming. Isoform 2/3 is likely to
CC interact with axonal proteins in close association with CNTNAP1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 1
CC colocalizes with ankyrin G at the nodes of Ranvier. Isoform 2/3 is

```

CC a glial component of the paranodal axo-glial junction.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1; Synonyms=NPI86, 186 kDa isoform;
 CC IsoId=P97685-1; Sequence=Displayed;
 CC Name=2; Synonyms=NPI155, 155 kDa isoform;
 CC IsoId=P97685-2; Sequence=VSP_050416, VSP_050417, VSP_050418,
 CC VSP_050419;
 CC Name=3; Synonyms=NPI155, 155 kDa isoform;
 CC IsoId=P97685-3; Sequence=VSP_050416, VSP_050417, VSP_008941,
 CC VSP_050418, VSP_050419;
 CC -1- TISSUE SPECIFICITY: Isoform 1 is expressed at Nodes of Ranvier
 CC while isoform 2/3 is expressed in unmyelinated axons.
 CC -1- DEVELOPMENTAL STAGE: Strongly but transiently up-regulated in
 CC oligodendrocytes at the onset of myelogenesis. Once these last
 CC have engaged their target axons, expression declines
 CC precipitously.
 CC -1- PTM: Isoform 2/3 is phosphorylated at p12. Dephosphorylation is
 CC required for ankyrin binding.
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
 CC LI/neurofascin/NG2CAM family.
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 4 fibronectin type III domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, AY016139; AAL27854.1; -;
 CC DR EMBL, U81035; AAB47753.1; -;
 CC DR HSRP, P20241; ICRB.
 CC DR InterPro; IPR003961; FN_III.
 CC DR InterPro; IPR008957; FN_III-like.
 CC DR InterPro; IPR007110; IG-like.
 CC DR InterPro; IPR003598; IG_c2.
 CC DR Pfam; PF00047; fn3; 4.
 CC DR Pfam; PF00047; Ig; 6.
 CC DR SMART; SM00060; FN3; 3.
 CC DR SMART; SM00408; IGC2; 4.
 CC DR PROSITE; PS50835; IG_Like; 6.
 CC DR Cell adhesion; Repeat; Signal: Transmembrane; Immunoglobulin domain;
 CC Glycoprotein; Alternative splicing; Phosphorylation.
 CC KW SIGNAL
 CC FT CHAIN 1 24
 CC FT 25 1240
 CC FT DOMAIN 25 1110
 CC FT TRAMEM 1111 1131
 CC FT DOMAIN 1132 1240
 CC FT DOMAIN 41 137
 CC FT DOMAIN 143 230
 CC FT DOMAIN 244 332
 CC FT DOMAIN 337 424
 CC FT DOMAIN 429 517
 CC FT DOMAIN 521 603
 CC FT DOMAIN 628 720
 CC FT DOMAIN 727 820
 CC FT DOMAIN 825 918
 CC FT DOMAIN 1008 1094
 CC FT DOMAIN 913 1006
 CC FT DISULFID 63 118
 CC FT DISULFID 162 213
 CC FT DISULFID 268 316
 CC FT DISULFID 358 408
 CC FT DISULFID 452 501
 CC FT DISULFID 543 592
 CC FT DISULFID 543 592
 CC FT CARBOHYD 305 305
 CC FT CARBOHYD 409 409
 CC FT CARBOHYD 446 446
 CC FT CARBOHYD 463 463

FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1015 1015 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1047 1047 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1101 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 31 36 Missing (in isoform 2 and isoform 3).
 FT VARSPLIC 236 236 T -> NNPTDSSLRNHPDIYSA (in isoform 2 and
 FT VARSPLIC 236 236 Isoform 3).
 FT VARSPLIC 611 625 Missing (in isoform 3).
 FT VARSPLIC 824 824 Missing (in isoform 3).
 FT VARSPLIC 928 1096 Missing (in isoform 2 and isoform 3).
 FT CONFLICT 482 482 E -> Q (IN REF. 2).
 FT CONFLICT 675 675 W -> L (IN REF. 2).
 FT CONFLICT 699 699 Q -> D (IN REF. 2).
 FT CONFLICT 763 763 Y -> YY (IN REF. 2).
 SQ SEQUENCE 1240 AA; 138003 MW; 616A187BC372513 CRC64;

Query Match 5.9%; Score 143.5; DB 1; Length 1240;
 Best Local Similarity 21.4%; Pred. No. 0.052;
 Matches 106; Conservative 67; Mismatches 193; Indels 129; Gaps 28;

QY 6 PFRLLVLQALALPAA-----TQGNKVVLGKKGPVELTCTASQ 45
 DB 8 PWTVAAILFLSLSGAIELPMDSIONELTPPTTKQSVKHIVDPNDIILECAKG 67
 QY 46 KKSIOFHW-KNSNQIKILNGQSFLLK-GPSKLNDRADSRSLMDQNPFLIKN---- 98
 DB 68 NPAPSPFWTNSRFFNIADKPRVSMRRSGTLVIDFSGGRPREY-EGEYQCFARNRFGT 126
 QY 99 -----LKIEDSTTYICEVEDQKEVQLVFGLTANSDTHLLOGGSLTTLBSPG-SS 150
 DB 127 ALSNRIRLQVSKSPW-----PKENLDPVV-----VDEGAPLTLQCNPPGLPS 170
 QY 151 PSVQCRSPRGNIQGGKTLIS-----VSQLELDSDSGT-WTCVTVLQN-----QKVEFK 196
 DB 171 PVTFMSSMEPTIQDKRVQSGHNGDLYFSNMQLDMQTDYSCNARHFTHTTQGNPFT 230
 QY 197 IDIVPCAPPEPKSCDKTHTCPPELLGDSVFLFPPEPKDTLMISRTPE--VTCVVVDVSHE 254
 DB 231 LKVLITRGVAERT-----PS-FMYPGQTSSQVLRGMILLLECTIASGV--- 273
 QY 255 DPEVKFMYVDGVEVHNAKTKPREQYNSTYRVSVTLVHQMNLNKEKCKKSNKALP 314
 DB 274 -PTPDIAWYKKGGLPSDKAK--FENFNKALRTINVEE-----DSGEYFCIASNMK- 322
 QY 315 APIEKTIS-KAKGQP---REPOVYTLPPSRDELTKQNVSLTCLVKGYPSDIAVEMESG 370
 DB 323 GSITHITSVYKAPFYLDPEKILILAPGEDG-----RLVCRANG-NPKP-TYQMLVNG 374
 QY 371 QPENNYKTPPVLDSD--GSFPLYSKLTVDKSRMQGNVSCSVMEALHNHYTKSL-- 426
 DB 375 DP---LQAPPNPBEVAGDTIIFRDQISSRA-----VYQCNSTNE--HGYYLANAFVS 424
 QY 427 -----SLSPGLQL 434
 DB 425 VLDPVPRMLSPRNQL 439
 RESULT 98
 KAC_MOUSE STANDARD; PRT; 106 AA.
 ID KAC_MOUSE
 AC P01837;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 02, Last annotation update)
 DE Ig kappa chain C region.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE (MOPC 21).
 RX MEDLINE=73053310; PubMed=4638343;
 RA Svasek J., Milstein C.;
 RT "The complete amino acid sequence of a mouse kappa light chain.";
 RL Biochem. J. 128:427-444(1972).
 RN (2)
 RP DISULFIDE BONDS (MOPC 21).
 RX MEDLINE=73008889; PubMed=5073237;
 RA Svasek J., Milstein C.;
 RT "The disulphide bridges of a mouse immunoglobulin G1 protein.";
 RL Biochem. J. 126:837-850(1972).
 RN (3)
 RP SEQUENCE FROM N.A., AND REVISIONS TO 53-59.
 RX MEDLINE=79084137; PubMed=103625;
 RA Hamlyn P.H., Brownlee G.G., Cheng C.-C., Galt M.J., Milstein C.;
 RT "Complete sequence of constant and 3' noncoding regions of an immunoglobulin mRNA using the diideoxynucleotide method of RNA sequencing.";
 RL Cell 15:1067-1075(1978).
 RN (4)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82059477; PubMed=6170937;
 RA Hamlyn P.H., Galt M.J., Milstein C.;
 RT "Complete sequence of an immunoglobulin mRNA using specific priming and the diideoxynucleotide method of RNA sequencing.";
 RL Nucleic Acids Res. 9:4485-4494(1981).
 RN (5)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81191915; PubMed=6262318;
 RA Max E.E., Maizel J.V. Jr., Leder P.;
 RT "The nucleotide sequence of a 5.5-kilobase DNA segment containing the mouse kappa immunoglobulin J and C region genes.";
 RL J. Biol. Chem. 256:5116-5120(1981).
 RN (6)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81198949; PubMed=6785724;
 RA Altenburger W., Neumaier P.S., Steinmetz M., Zachau H.G.;
 RT "DNA sequence of the constant gene region of the mouse immunoglobulin kappa chain.";
 RL Nucleic Acids Res. 9:971-981(1981).
 RN (7)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88329081; PubMed=3181116;
 RA de Waele P., Feys V., van de Voorde A., Molemans F., Piers W.;
 RT "Expression in non-lymphoid cells of mouse recombinant immunoglobulin directed against the tumour marker human placental alkaline phosphatase.";
 RL Eur. J. Biochem. 176:287-295(1988).
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: V00807; CAA24169.1; -
 DR PIR: B90262; KIMS.
 DR PDB: 1AIF; 01-FEB-97.
 DR PDB: 1FSK; 02-OCT-00.
 DR PDB: 1KBS; 08-APR-98.
 DR PDB: 1KCR; 11-MAY-02.

DR PDB: 1KCS; 11-MAY-02.
 DR PDB: 1KCU; 11-MAY-02.
 DR PDB: 1KCV; 11-MAY-02.
 DR PDB: 25C8; 09-JUL-99.
 DR MGD; MGI:96495; Igk-C.
 DR InterPro; IPR007110; Ig-1like.
 DR InterPro; IPR003597; Ig_C1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig_1.
 DR SMART; SM00407; IG1; 1.
 DR PROSITE; PS50835; IG-LIKE; 1.
 DR PROSITE; PS00290; IG_MHC; 1.
 KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
 FT NON_TER 1 102 IG-LIKE.
 FT DOMAIN 5 102 IG-LIKE.
 FT DISULFID 26 86
 FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
 SQ SEQUENCE 106 AA; 11778 MW; 4851PFSEF49BAEBS CRC64;
 Query Match 5.9%; Score 143; DB: 1; Length 106;
 Best Local Similarity 31.5%; Pred. No. 0.0026;
 Matches 29; Conservative 18; Mismatches 33; Indels 12; Gaps 2;
 QY 331 PNYVTLPPSRDELTKNQSILTCVKGFPSPDIAVEESNGQPENNYKTPPYLDS----- 385
 DB 5 PTVSIFPPSSSEGLTSGASVCFLNNFYPKQINVKMKDKDSERQN-----GVLSNWTQD 59
 QY 386 --DGSFFLYSKLTVDKSRMQGNVFGSGVME 415
 DB 60 SKDSTYSMSSTLTLLKDEYERHNSYTCRATHK 91
 RESULT 99
 ID CAML_MOUSE STANDARD; PRT; 1260 AA.
 AC P11627;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neutral adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin.
 GN L1CAM OR CAML1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=88318924; PubMed=3412448;
 RA Moos M., Tacke R., Scherer H., Teplov D., Fries K., Schachner M.;
 RT "Neutral adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin.";
 RL Nature 334:701-703(1988).
 CC -1- FUNCTION: Cell adhesion molecule with an important role in the development of the nervous system. Involved in neuron-neuron adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds to axonin on neurons.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
 CC L1/neurofascin/NG2CAM family.
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: X12875; CAA31368.1; -
 DR PIR: S05479; S05479.

```
DR HSP; P20241; 1CFB.
DR MGD; MGI:96721; L1cam.
DR GO; GO:0007411; P:axon guidance; IMP.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00047; fn3; 6.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 5.
DR PROSITE; PS50835; IG_Like; 6.
DR Neogenesis; Cell adhesion; Developmental protein; glycoprotein;
KW Transmembrane; Repeat; Immunoglobulin domain; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1124 1146 POTENTIAL.
FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 130 IG-LIKE C2-TYPE 1.
FT DOMAIN 138 225 IG-LIKE C2-TYPE 2.
FT DOMAIN 239 327 IG-LIKE C2-TYPE 3.
FT DOMAIN 332 419 IG-LIKE C2-TYPE 4.
FT DOMAIN 424 506 IG-LIKE C2-TYPE 5.
FT DOMAIN 517 600 IG-LIKE C2-TYPE 6.
FT DOMAIN 827 896 FIBRONECTIN TYPE-III 1.
FT DOMAIN 932 994 FIBRONECTIN TYPE-III 2.
FT DOMAIN 1032 1094 FIBRONECTIN TYPE-III 3.
FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 552 564 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 57 113 BY SIMILARITY.
FT DISULFID 157 208 BY SIMILARITY.
FT DISULFID 263 311 BY SIMILARITY.
FT DISULFID 353 403 BY SIMILARITY.
FT DISULFID 447 496 BY SIMILARITY.
FT DISULFID 538 590 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1260 AA; 140968 MW; 22B557001CB2A538 CRC64;

Query Match 5.9%; Score 143; DB 1; Length 1260;
Best Local Similarity 22.1%; Pred. No. 0.057;
Matches 89; Conservative 53; Mismatches 141; Indels 120; Gaps 20;
```

```
Db 609 GSPGPVPHLELSDRHLLKQSQVHL.SW-----SPAEDHNSPIEKYDIEFDKENAPKMF 662
Oy 170 SVSQLELDGSGTWTCTVLAQNKVEFKIDIVPC-----PADEPKSCDHTHCPBLLGG 222
Db 663 SLGKV-----FQNTSTLLKSLPYVHYFRYTAINKVGPSPVSESVTPPEAPF----- 714
Oy 223 PSVFLLPPKPRDITLMSRTPEVTCVVVDVSHEDPEVK-----FMWYDGEVHN 271
Db 715 -----KNP-----VDVAGEGNETNNMVIYTWKPLRMMDMNAPOIQ-YR 750
Oy 272 AKTKPREQYNSYTRVVSULTVLHQMVLNGEKYKCKVSNKALPAPLEKTSKAKGPR-- 329
Db 751 VQMPVQCKQ--ETWRKQTV-----SDPLVSNSTVTPVEIKQAVANNQCKGP 797
Oy 330 EPQV---YT-----LPPSRDELTKQVSLTCLVKGFPSDIA 363
Db 798 EPQVTTIGSGEDYRQVSPLEEDITFN-SSTVLVR-WRPVDLA 838

RESULT 100
ID_NFAS_HUMAN STANDARD; PRT; 1240 AA.
AC O94855; O96K50.
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Neurofascin precursor.
GN NFASC OR KIA0756.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 866-1240 (ISOFORM 1), AND
SEQUENCE OF 804-1240 FROM N.A. (ISOFORM 4).
RP RP
RC RP
RA KAWAKAMI B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Matsushima M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 31-1240 FROM N.A. (ISOFORM 3).
RP RP
RC RP
RA MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
[3]
SEQUENCE OF 326-1240 FROM N.A. (ISOFORM 2).
RP RP
RC RP
RA MEDLINE=22388257; PubMed=12477932;
RA TISSUE=Brain;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Ditschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant P.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.T.,
RA Rana S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McWay P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Vallatou D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
```

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marx M.A.:
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences";
CC -I- FUNCTION: Cell adhesion, ankyrin-binding protein which may be
CC involved in neurite extension, axonal guidance, synaptogenesis,
CC myelination and neuron-glia cell interactions (By similarity).
CC -I- SUBUNIT: Probable constituent of a neurofascin/NRCAM/ankyrin G
CC complex. Associates with the sodium channel beta-1 (SCN1B) and
CC beta-3 (SCN3B) subunits (By similarity).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1:
CC IsoId=O94856-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=2;
CC IsoId=O94856-2; Sequence=VSP_008937, VSP_008938;
CC Name=3;
CC IsoId=O94856-3; Sequence=VSP_008939;
CC Note=No experimental confirmation available;
CC Name=4;
CC IsoId=O94856-4; Sequence=VSP_008940;
CC Note=No experimental confirmation available;
CC -I- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC -I- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -I- SIMILARITY: Contains 4 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sb.ch).
CC -----
DR EMBL AK027553 BAB55195.1; -;
DR EMBL AK090639 -; NOT_ANNOTATED_CDS.
DR EMBL AK127424 -; NOT_ANNOTATED_CDS.
DR EMBL AB018299 BAA34476.2; -;
DR EMBL BC008124 AAH08124.2; -;
DR HSSP P20241 1CFB.
DR InterPro IPR003961; FN_III.
DR InterPro IPR008957; FN_III-like.
DR InterPro IPR003599; Ig.
DR InterPro IPR007110; Ig-like.
DR InterPro IPR003598; Ig_c2.
DR Pfam PF00047; Ig_6.
DR Pfam PF00060; FN3; 3.
DR SMART SM00409; IGc2; 6.
DR SMART SM00408; IGc2; 6.
DR PROSITE PS50835; IG_Like; 6.
DR Cell adhesion. Repeat; Signal; Transmembrane; Immunoglobulin domain;
KW Glycoprotein; Alternative splicing; Polymorphism.
KW
FT CHAIN 1 24
FT SIGNAL 1 24
FT CHAIN 25 1240
FT SIGNAL 25 1240
FT DOMAIN 25 1110
FT TRAMEM 1111 1131
FT DOMAIN 1132 1240
FT DOMAIN 41 137
FT DOMAIN 143 230
FT DOMAIN 244 332
FT DOMAIN 337 424
FT DOMAIN 429 517
FT DOMAIN 521 603
FT DOMAIN 628 720
FT DOMAIN 727 820
FT DOMAIN 825 924
FT DOMAIN 1008 1094
FT DOMAIN 924 1001

FT DISULFID 63 118 POTENTIAL.
FT DISULFID 162 213 POTENTIAL.
FT DISULFID 268 316 POTENTIAL.
FT DISULFID 358 408 POTENTIAL.
FT DISULFID 452 501 POTENTIAL.
FT DISULFID 543 592 POTENTIAL.
FT CARBOHYD 305 305 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 446 446 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1015 1015 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1026 1026 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1047 1047 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT VARSPLIC 611 619 ADOARPTNR -> GNCPCSPWH (in isoform 2).
FT VARSPLIC 620 1071 Missing (in isoform 2).
FT VARSPLIC 928 1096 Missing (in isoform 3).
FT VARSPLIC 923 1096 Missing (in isoform 4).
FT VARIANT 159 159 T -> M (in dbSNP:3795564).
FT CONFLICT 807 807 F -> L (in REF:1; BAB55195).
SQ SEQUENCE 1240 AA; 137618 MW; 2A89A9D7CD28596 CRC64;

Query Match 5.9%; Score 142.5; DB 1; Length 1240;
Best Local Similarity 21.7%; Pred. No. 0.061;
Matches 97; Conservative 64; Mismatches 176; Indels 109; Gaps 27;
QY 35 DTVELTCTASOKSKSIQIFHW-KNSNQIKILGNQSFYLR-GPSKLNDRADRRSLMDQGN 91
Db 57 DNILIECPAKGNPASFMTNRSRPFNIAXDPVSRMRSGTLVIDFRSGGRPEBY-RGE 115
QY 92 FPLIYKN-----LKIEDSPYICVEVDQKEBYQLLVFGLTANSDFHLOGSLT 140
Db 116 YOCFARNFEGTALSNRILQVSKSPLW-----PKENLDPVV-----VOEGAPLT 159
QY 141 LTLSEPPG-SSPSVQCRSPRGNIQGGKTLG-----VSOLELDQSGT-WTCVTLQN- 189
Db 160 LQCNPPGLPSFVIRFMWSSMBEPIQODKRVQGHGDIYFSNWLQDMQTYSCARHFF 219
QY 190 ----QKVEFKIDIVPCPAPPEKSCDKTHTCPELLGSPVFLFPKPKDTLMISRTPE-- 243
Db 220 THTIQKXNPFTLKVLTRGVAERT-----PS-FMYPQGTASQGVLRGMDDL 265
QY 244 VTCVVVDVSHEDPEVKFMYVDGVEFVNAAKTKPREEQNSTYRVVSVLTVLHQMNLNKE 303
Db 266 LECTASGV-----PPFDIAMYKKGGLPSDKAK--FENFNKALRTITVSEB-----DSGE 313
QY 304 YKCKVSNLALPAPIEKTIS-KAKGQ---REPOVYTLPPSDELTQNQVSLTCLVKGFYP 359
Db 314 YFCIASNM--GSIHTTISVRKAAFYWLDEBKILILPGEDEG-----RLVCRANG--NP 364
QY 360 SDIAVEWESNQCPENNNYKTPPVLDSD--GSFFLYSKLTVDKSRNQGNVSCSYMHBAL 417
Db 365 KP-TQVMWNGEP---LQSAFPNPRREVAAGDTIIFRDQISRA-----VYQNTSNB-- 413
QY 418 HNHYTQKSL-----SLSPGLQL 434
Db 414 HGYLANAFAVSLVDVPRMLSPRNL 439

Search completed: August 3, 2004, 13:09:06
Job time : 12.2349 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:51:44 ; Search time 31.1743 Seconds
(without alignments)
4594.975 Million cell updates/sec

Title: SEQ4
Perfect score: 2414
Sequence: 1 MNRGVFPHLLVLQLALLP.....DETCAADQGEUDGJMTTDP 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriaph: *
17: SP archaeop: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1273	52.7	470	4	Q725W1
2	1260.5	52.2	482	4	Q72351
3	1255.5	52.0	469	4	Q727P5
4	1252	51.9	509	4	Q8NF17
5	1213	50.2	679	4	Q96P08
6	1180.5	48.9	521	4	Q8NKY9
7	1156.5	47.9	354	4	Q86T72
8	1151	47.7	437	4	Q8T6C3
9	912	37.8	337	6	Q95M34
10	884	36.6	430	11	Q7TMR1
11	867.5	35.9	437	11	Q9R1A4
12	858.5	35.6	469	11	Q8R3V9
13	848.5	35.1	463	11	Q99LCA
14	846	35.0	468	11	Q99L31
15	833.5	34.5	473	11	Q9D8L4
16	823.5	34.1	473	11	Q99L25

17	809.5	33.5	473	11	Q91205	Q91205 mus musculus
18	808.5	33.5	474	11	Q8R3H6	Q8R3H6 mus musculus
19	767	31.8	397	6	Q09261	Q09261 cercopithec
20	767	31.8	397	6	Q09262	Q09262 cercopithec
21	767	31.8	397	6	Q09263	Q09263 cercopithec
22	767	31.8	397	6	Q09269	Q09269 cercopithec
23	763	31.6	397	6	Q09259	Q09259 cercopithec
24	763	31.6	397	6	Q09260	Q09260 cercopithec
25	758	31.4	397	6	Q092805	Q092805 cercopithec
26	743	30.7	457	6	Q8H2T8	Q8H2T8 callithrix
27	740.5	30.7	457	6	Q8H2T7	Q8H2T7 saimiri sci
28	601.5	24.9	455	6	Q9XS78	Q9XS78 delphinapne
29	557.5	23.1	474	6	P79355	P79355 felis silve
30	467	19.3	457	11	Q61396	Q61396 mus musculus
31	436.5	18.1	433	11	Q55054	Q55054 mus musculus
32	400	16.6	86	6	Q77596	Q77596 mandrillus
33	400	16.6	86	6	Q77597	Q77597 mandrillus
34	397	16.4	86	6	Q77594	Q77594 cercopithec
35	397	16.4	86	6	Q77599	Q77599 theropithec
36	395.5	16.4	613	4	Q96EY0	Q96EY0 homo sapien
37	395	16.4	86	6	Q77595	Q77595 cercocobus
38	392	16.2	86	6	Q77598	Q77598 papio sp. (
39	389.5	16.1	613	4	Q8WUK1	Q8WUK1 homo sapien
40	389.5	16.1	614	4	Q96GA6	Q96GA6 homo sapien
41	389.5	16.1	618	4	Q96AA6	Q96AA6 homo sapien
42	388	16.1	86	6	Q77601	Q77601 lophocobus
43	385	15.9	597	4	Q9BU10	Q9BU10 homo sapien
44	385	15.9	597	4	Q9BQ88	Q9BQ88 homo sapien
45	385	15.9	597	4	Q96BB9	Q96BB9 homo sapien
46	383	15.9	86	6	Q77600	Q77600 lophocobus
47	381	15.8	614	11	Q7TMT6	Q7TMT6 mus musculus
48	380	15.7	588	4	Q8WUX4	Q8WUX4 homo sapien
49	372.5	15.5	613	11	Q8VCX7	Q8VCX7 mus musculus
50	366.5	15.2	375	4	Q86T11	Q86T11 homo sapien
51	357	14.8	71	4	Q13969	Q13969 homo sapien
52	345.5	14.3	587	13	Q7TOR1	Q7TOR1 xenopus lae
53	336	13.9	478	4	Q72379	Q72379 homo sapien
54	332.5	13.4	492	4	Q72374	Q72374 homo sapien
55	315.5	13.1	484	11	Q99L66	Q99L66 mus musculus
56	314.5	13.0	487	11	Q99KA4	Q99KA4 mus musculus
57	313.5	12.9	489	11	Q8VCX4	Q8VCX4 mus musculus
58	312	12.9	416	4	Q9NPP6	Q9NPP6 homo sapien
59	311	12.9	481	11	Q91WT1	Q91WT1 mus musculus
60	311	12.9	482	11	Q8K172	Q8K172 mus musculus
61	307	12.7	482	11	Q81X92	Q81X92 mus musculus
62	305.5	12.6	99	6	Q29027	Q29027 sus scrofa
63	305	12.6	486	11	Q91207	Q91207 mus musculus
64	304	12.6	488	11	Q8K0F2	Q8K0F2 mus musculus
65	301.5	12.5	684	13	Q90544	Q90544 ginglymosto
66	300.5	12.4	480	11	Q91XR1	Q91XR1 mus musculus
67	300.5	12.4	494	4	Q96K68	Q96K68 homo sapien
68	298.5	12.4	481	11	Q91WT3	Q91WT3 mus musculus
69	297.5	12.3	484	11	Q8VEA0	Q8VEA0 mus musculus
70	297.5	12.3	487	11	Q80Z17	Q80Z17 mus musculus
71	297.5	12.3	496	4	Q96KX8	Q96KX8 homo sapien
72	295	12.2	493	4	Q8NCL6	Q8NCL6 homo sapien
73	293	12.1	488	11	Q91WR1	Q91WR1 mus musculus
74	290	12.0	499	4	Q8NSK4	Q8NSK4 homo sapien
75	287.5	11.9	496	4	Q96DK0	Q96DK0 homo sapien
76	286.5	11.9	426	11	Q9DCD9	Q9DCD9 mus musculus
77	286.5	11.9	479	11	Q91WP5	Q91WP5 mus musculus
78	284.5	11.8	479	11	Q7TMR4	Q7TMR4 mus musculus
79	284.5	11.8	481	11	Q8VCY5	Q8VCY5 mus musculus
80	284	11.8	497	4	Q8WY24	Q8WY24 homo sapien
81	281.5	11.7	384	4	Q9UP60	Q9UP60 homo sapien
82	281	11.6	479	11	Q99M22	Q99M22 mus musculus
83	280.5	11.6	99	6	Q29028	Q29028 sus scrofa
84	273.5	11.3	500	4	Q9BRV0	Q9BRV0 homo sapien
85	270	11.2	480	11	Q8K0Z4	Q8K0Z4 mus musculus
86	189	7.8	482	13	Q90WB5	Q90WB5 anas platyr
87	186	7.7	268	13	Q90524	Q90524 ginglymosto
88	183.5	7.6	130	11	Q9D8W4	Q9D8W4 mus musculus
89	183.5	7.6	234	4	Q8N355	Q8N355 homo sapien

```

90 183.5 7.6 235 11 Q99M11 Q99M11 mus musculus
91 182 7.5 236 4 Q7Z3Y4 Q7Z3Y4 homo sapien
92 181.5 7.5 955 4 Q8NFP4 Q8NFP4 homo sapien
93 181 7.5 379 11 Q8BIX5 Q8BIX5 mus musculus
94 181 7.5 397 11 Q8BFX8 Q8BFX8 mus musculus
95 179.5 7.4 573 4 Q8WJ38 Q8WJ38 homo sapien
96 176.5 7.3 233 4 Q96I69 Q96I69 homo sapien
97 176.5 7.3 259 13 Q90510 Q90510 ginglymosto
98 176.5 7.3 739 6 Q28260 Q28260 canis famli
99 176.5 7.3 739 11 Q63669 Q63669 rattus norv
100 174 7.2 233 4 Q8N5P4 Q8N5P4 homo sapien
101 174 7.2 236 4 Q96E61 Q96E61 homo sapien
102 173 7.2 233 4 Q8TBC9 Q8TBC9 homo sapien
103 173 7.2 234 4 Q72473 Q72473 homo sapien
104 172.5 7.1 234 4 Q722U7 Q722U7 homo sapien
105 171.5 7.1 257 13 Q90536 Q90536 ginglymosto
106 171.5 7.1 885 6 Q8HYV1 Q8HYV1 sus scrofa
107 171.5 7.1 886 6 Q8HYV2 Q8HYV2 sus scrofa
108 170 7.0 237 13 Q90545 Q90545 ginglymosto
109 170 7.0 238 7 Q9WXA2 Q9WXA2 autonocara
110 168.5 7.0 811 5 Q9VNP2 Q9VNP2 drosophila
111 168.5 7.0 1011 5 Q24273 Q24273 drosophila
112 167.5 6.9 267 13 Q90529 Q90529 ginglymosto
113 167.5 6.9 739 6 Q9GKR3 Q9GKR3 bos taurus
114 167 6.9 105 11 Q99JC1 Q99JC1 mus musculus
115 166 6.9 240 4 Q8NEK0 Q8NEK0 homo sapien
116 166 6.9 239 4 Q8WUK3 Q8WUK3 homo sapien
117 165 6.8 106 4 Q8TCJ5 Q8TCJ5 homo sapien
118 165 6.8 237 4 Q8WUK4 Q8WUK4 homo sapien
119 165 6.8 237 4 Q8WTU6 Q8WTU6 homo sapien
120 165 6.8 650 6 Q9GKR2 Q9GKR2 bos taurus
121 164 6.8 8081 5 Q72120 Q72120 caenorhadi
122 163.5 6.8 584 4 Q9Y3Y8 Q9Y3Y8 homo sapien
123 162.5 6.7 252 13 Q90568 Q90568 ginglymosto
124 162.5 6.7 2828 4 Q9NR99 Q9NR99 homo sapien
125 161 6.7 259 7 Q8SPB4 Q8SPB4 sus scrofa

```

ALIGNMENTS

```

RESULT 1
Q7Z5W1 PRELIMINARY; PRT; 470 AA.
AC Q7Z5W1;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxId=9606;
RX MEDLINE=223828257; PubMed=12477932;
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheiner C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshynski S., Carlini P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeck S.A., McGowan P.J., McKernan K.O., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villard D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,

```

```

RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP [2]
RC TISSUE=Spleen;
RA Straube R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053984; AAH53984.1; -.
KM Hypothetical protein.
SQ SEQUENCE 470 AA; 51204 MW; 778CF34521483E1A CRC64;

```

Query Match 52.7%; Score 1273; DB 4; Length 470;

Best Local Similarity 59.3%; Pred. No. 4.2e-98;

Matches 274; Conservative 25; Mismatches 81; Indels 82; Gaps 11;

```

QY 30 LGKKDITVELTCTAS--QKSIQPHW-----KNSNOIKILNGQSFLLTGPSK--LND 78
DB 30 LVQPGSLRLSCVASGFTLNNDYDMHWYQGIKGLKLEWVSXIGTADRYVAGSVYGRFTIS 89
QY 79 RASRRSLMDQGNPPLIKNLKIEDSDTYICEVEDQKEVQVLVFGILTASDTHLQGS 138
DB 90 RENAQSGLYQNM-----SLRVDAAVYVC-----ARGAKWAPLGAFLDWGG 133
QY 139 LTTLESPPSSPSVQCSRGRKNIQGG-----KTLVS----- 172
DB 134 TMTVVSASATKGPVFPFLAPSSKTSSTGTAALGLVDYFPEPYTVSWNSGALTSVHTF 193
QY 173 QLEIQDSG-----TWCTVLONOKQKFEKIDYPCAPPEKSCKTH 214
DB 194 PAVLQSSGLYLSVTVTPSSSLGTQYICNV--NHRPSNTKVD---KVEPKSCDKTH 247
QY 215 TC-----PELLGSPVFLPPKPKDITMISRTPEVTCVVVDVSHEDPEVKFNVVDQEV 269
DB 248 TCEPCAPPELLGGSVFLFPPKPKDITMISRTPEVTCVVVDVSHEDPEVKFNVVDQEV 307
QY 270 HNAKTPREQYNSTYRVSVLTITVLHODMNLGKRYKCKVSNKALPAPIETKISKAKQPR 329
DB 308 HNAKTPREQYNSTYRVSVLTITVLHODMNLGKRYKCKVSNKALPAPIETKISKAKQPR 367
QY 330 EPQVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVEMESNGQBPNNYKTPPVLDSDGSF 389
DB 368 EPQVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVEMESNGQBPNNYKTPPVLDSDGSF 427
QY 390 FLVSKLTVDSKRMQGVFSCVWHEALHNHYOKSLSLSPG 431
DB 428 FLVSKLTVDSKRMQGVFSCVWHEALHNHYOKSLSLSPG 469

```

RESULT 2

```

Q7Z351 PRELIMINARY; PRT; 482 AA.
AC Q7Z351;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZP686N02209.
GN DKFZP686N02209.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxId=9606;
RX MEDLINE=223828257; PubMed=12477932;
RP SEQUENCE FROM N.A.
RC TISSUE=Human rectum tumor;
RA Bloecher H., Boecker M., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX538118; CAD98026.1; -.
KM Hypothetical protein.
SQ SEQUENCE 482 AA; 52852 MW; EDAV5FI901DIA034 CRC64;

```



```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Primary B-Cells;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033178; ANH33178.1; -.
DR PIR; A60764; A60764.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_w.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; Igcl; 3.
DR SMART; SM00406; IgV; 1.
DR PROSITE; PS50835; IG_Like; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein
SQ
SEQUENCE 521 AA; 57156 MW; 2AC7D22E72D6CAA2 CRC64;

Query Match          48.9%; Score 1180.5; DB 4; Length 521;
Best Local Similarity 51.5%; Pred. No. 2.7e-90;
Matches 261; Conservative 40; Mismatches 85; Indels 121; Gaps 13;

QY 30 LGKGDVTELTCTAGSQKSIQFMKNSNQIKILGNQ--GSPITKGPDKLND----- 78
DB 30 LVQPGSGSLRLSCAASGFIVSDHYEWNVRCAPGKGPWVCSPFRSKHAKSTTEEAASVGRF 89
QY 79 ----RADSRRLWDQGNFLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHL-L 134
DB 90 TILRDDSKNSSVHLQNN-----SLKTDYDAVYYC-VRDLEG-----AGKIDWYEDI 133
QY 135 OGQSLLTLESPPGSSPSVQCRSPRGKNIQGG-----KTLSVS----- 172
DB 134 WGRGLIVTVSSASTKGPSVFLPAPCSRSTSGGTAALGCLVKDYFPEPYTVSNNSGALTSG 193
QY 173 ----QLELDDSG-----TWICTV-----LONOKVVEFKIDI----- 199
DB 194 VHTFPAVLQSSGLYSLSVTVTPSSSLGTQTYTCVNNHKSNTKYDKRVELKTPRLGDTTH 253
QY 200 -----VPCP-APEPKSCDKHTHC-----PELLGGS 224
DB 254 TPCRCPEPKSCDTPPCPCPEPKSCDTPPCPCPEPKSCDTPPCPCPAPELLGGS 313
QY 225 VFLFPPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNST 284
DB 314 VFLFPPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNST 373
QY 285 YRVSVSLTVLHODMLNGKEYKCKVSKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTL 344
DB 374 FRVSVSLTVLHODMLNGKEYKCKVSKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTL 433
QY 345 KRVQVSLTCLVKGFPYSDIIVEMESNGQPENNYKTTPPVLDSDGSFELYSKLTVDKSRNQ 404
DB 434 KRVQVSLTCLVKGFPYSDIIVEMESNGQPENNYKTTPPVLDSDGSFELYSKLTVDKSRNQ 493
QY 405 GNVFSCSVMEALAHNYTKSLSPG 431
DB 494 GNVFSCSVMEALAHNYTKSLSPG 520

RESULT 7
Q86TT2 PRELIMINARY; PRT; 354 AA.
AC Q86TT2;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Human full-length cDNA clone CS0D1019YF20 of placenta of Homo sapiens
DE (Human) (Fragment).
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Genoscope;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX248278; CAD62606.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; Igcl; 3.
DR SMART; SM00406; IgV; 1.
DR PROSITE; PS50835; IG_Like; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Plasmid
FT NON_TER
SQ
SEQUENCE 354 AA; 39125 MW; 23B80BF4D2B87A92 CRC64;

Query Match          47.9%; Score 1156.5; DB 4; Length 354;
Best Local Similarity 61.2%; Pred. No. 1.6e-88;
Matches 232; Conservative 25; Mismatches 45; Indels 77; Gaps 9;

QY 87 WDOG-----NPLIINKLKIED-----SDTYICEVE-----DQKEEV 118
DB 18 WNSGALTSQGVHTFPAVLQSSGLYSLSVTVTPSSSLGTQTYTCVNNHKSNTKYDKRVEL 77
QY 119 QLVGLVTRANSPTHLLOQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLD 178
DB 78 K-----TVLGDTHTCPCPEPKSCDTPP-----PCPCPEPK----- 110
QY 179 SGTWTCVTLQONOKVVEFKIDIVPCP-APEPKSCDKHTHC-----PELLGGSVFLFPPKP 232
DB 111 -----SCDTPP-----PCRCPEPKSCDTPPCPCPEPKSCDTPPCPCPAPELLGGSVFLFPPKP 154
QY 223 KDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTFRVSVLT 292
DB 155 KDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTFRVSVLT 214
QY 223 VILHODMLNGKEYKCKVSKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTLKNQVSLTC 352
DB 215 VILHODMLNGKEYKCKVSKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTLKNQVSLTC 274
QY 353 LVKGFYSPDIDIVEMESNGQPENNYKTTPPVLDSDGSFELYSKLTVDKSRMQGNVFSGSV 412
DB 275 LVKGFYSPDIDIVEMESNGQPENNYKTTPPVLDSDGSFELYSKLTVDKSRMQGNVFSGSV 334
QY 413 MHEALAHNYTKSLSPG 431
DB 335 MHEALAHNYTKSLSPG 353

RESULT 8
Q8TC63 PRELIMINARY; PRT; 473 AA.
AC Q8TC63;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
```

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strauberg R;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC025985; AAH25985.1; -
 DR GO; GO:0005507; F: copper ion binding; IEA.
 DR GO; GO:0005489; F: electron transporter activity; IEA.
 DR GO; GO:0006118; F: electron transport; IEA.
 DR InterPro; IPR000923; BlueCu.1.
 DR InterPro; IPR007110; Ig-1-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS00196; COPPER BLUE; 1.
 DR PROSITE; PS00835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 DR Hypothetical protein.
 SQ SEQUENCE 473 AA; 51986 MW; E29920B09BA369F5 CRC64;

Query Match 47.7%; Score 1151; DB 4; Length 473;
 Best Local Similarity 53.7%; Pred. No. 7e-88;
 Matches 262; Conservative 33; Mismatches 105; Indels 88; Gaps 16;

QY 8 RHLLVLQLALP-----AATQGNKVVIGKKGDTVELTCTAS--QKSKIQFM----- 53
 DB 9 KHLWFLLLVAAPRVNLSRLQDSEGPLKPSVTLSTLCYVSGSVASSTYMGWRQP 68
 QY 54 --KNSNQIKILGNQSPFLTKGSPKLNDRADSRRLMDQGNPPLIIKNIKIEDSDTYICEV 111
 DB 69 PCKGLEWIGTINFSGN--MYSPS--LRSRYTMSADM--SENSFYLKUDSVTAADTAVYCAA 125
 QY 112 EDQKEVQLVLFGLTANSDTHLQGSILTLIESPPSS--PSVQCRPRPKNI----- 163
 DB 126 GH-----LVWGFGA-----HMGGKLVSVSPASTKGPVFPPLAPCSRSTSESTALGCL 174
 QY 164 -----OGCKTLVSQLE--LQDSG-----TWICTV--- 186
 DB 175 VKDYREPVTVSWNSGALTSQVHTFPAVLQSSGLXLSVTVTPSSISGTYTCNVNHHK 234
 QY 187 ---LONQKVEPKLIDVPCPAPEPKSCQKTHTCPELLGSPVLFPPPKXDTLMSRTP 243
 DB 235 PSNTKVDKRVESKYG--PCPC-----SC-----PAPEFLGAPSVLFPPPKXDTLMSRTP 284
 QY 244 VTCVVVDVSHDEPVKEMVYDGVENNAKTKPREQYNSTYRVVSVLTVLHQDLNKE 303
 DB 285 VTCVVVDVSDPEVQFMVYDGVENNAKTKPREQFNSTYRVVSVLTVLHQDLNKE 344
 QY 304 YKCKVSNKALPAPIEKTSKAKGQPREPOVYTLPPSRDELTKNQVSTCLVKGFYPSDIA 363
 DB 345 YKCKVSNKGLPSSIEKTSKAKGQPREPOVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIA 404
 QY 364 VEWESNGQPENNYKTTPPVLDSDGSFPLYSKLTVDKSRMQQGNVSCVMHEALHNYHQ 423
 DB 405 VEWESNGQPENNYKTTPPVLDSDGSFPLYSKLTVDKSRMQQGNVSCVMHEALHNYHQ 464
 QY 424 KSLISLSPG 431
 DB 465 KSLISLSPG 472

RESULT 9
 Q9SM34 PRELIMINARY; PRT; 337 AA.
 AC Q9SM34;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Immunoglobulin gamma 1 heavy chain constant region (Fragment).
 GN IGH1.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wagner B.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98383416; PubMed=9717671;
 RA Wagner B., Overesch G., Sheoran A., Holmes M., Richards C.,
 Leibold W., Radduch A.;
 RT "Organization of the equine immunoglobulin heavy chain constant region
 genes. III. Alignment of c-mu, c-gamma, c-epsilon and c-alpha genes.";
 RL Immunobiology 199;105-119(1998).
 DR EMBL; AJ306675; CAC44624.1; -
 DR InterPro; IPR007110; Ig-1-like.
 DR InterPro; IPR003597; Ig_C1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 2.
 DR SMART; SM00407; IGc1; 2.
 DR PROSITE; PS00835; IG_LIKE; 3.
 DR PROSITE; PS00290; IG_MHC; 2.
 DR NON_TER 1
 SQ SEQUENCE 337 AA; 37438 MW; A60BF2B01DEFD1F6 CRC64;

Query Match 37.8%; Score 912; DB 6; Length 337;
 Best Local Similarity 59.1%; Pred. No. 4.6e-68;
 Matches 179; Conservative 40; Mismatches 60; Indels 24; Gaps 7;

QY 142 TLSEPPSSPSVOCSPRGNKIQGKTLVSQLELDPSGTWT-----CTVLONQKVEFK 196
 DB 45 SLTSGVHTFSPSV-----LQSGFYSLSMVTVASTWTSETYICNVVHAAS--NFK 93
 QY 197 ID--IYVCPAPEKSCD--KTHTC--PELLGSSVLFPPPKXDTLMSRTPVTVND 250
 DB 94 VDKRIEIPIDNHOKVCMSCPCPCPAPELLGSSVLFPPPKXDTLMSRTPVTVND 153
 QY 251 VSHDEPVKEMVYDGVENNAKTKPREQYNSTYRVVSVLTVLHQDLNKEKCKVSN 310
 DB 154 VSGENPVKEMVYDGVENNAKTKPREQFNSTYRVVSVLTVLHQDLNKEKCKVSN 213
 QY 311 KALPAPIEKTSKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNG 370
 DB 214 QALPAPIEKTSKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNG 273
 QY 371 QP--ENNYKTTPPVLDSDGSFPLYSKLTVDKSRMQQGNVSCVMHEALHNYHQKSL 428
 DB 274 QPELETYSTTQAQODSDGSFPLYSKLTVDKSRMQQGTPTCGVMHEALHNYHQKSVK 333
 QY 429 SPG 431
 DB 334 NPG 336

RESULT 10
 Q7TMK1 PRELIMINARY; PRT; 470 AA.
 AC Q7TMK1;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SMPAIN=CZECH II; TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
 RA Dlachenko L., Marinisa K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedini T.B., Tothlyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Breast tumor;
 RA Straubeberg R.,
 RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC055910; AAH55910.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 470 AA; 51727 MW; 6D90E4DF896B090 CRC64;
 Query Match 36.6%; Score 884; DB 11; Length 470;
 Best Local Similarity 43.0%; Pred. No. 1.6e-65;
 Matches 196; Conservative 61; Mismatches 129; Indels 70; Gaps 12;
 Oy 30 LGKGGDTVELTCTAS--QKSIQFHWKNSNQIKILNGOSF-LTKGPKSLNDPADRSRL 86
 Db 30 LKPGASVATSCAASGYTGTGYMHWVSKSHGSLWGLVNPBNDGTITNKGFKKATL 89
 Oy 87 W---DGNFPLIKNIKIEDSDTYICEVEDQKEVOLVFGLTANSDTHLQGSLLTL 143
 Db 90 TVDKSSSTAYVMEINSLTSEDSAYVC-----AAYYSGSYWYD--VWGAGTTTV 138
 Oy 144 ESPGSSPSVQ-----CRSPRGKNT-----QGKTL----- 169
 Db 139 SASATTAPSVPLVPCGCDTSGSSVTGLGVKGYFPEPYTVKNNGALSSGVATYSSVLQ 198
 Oy 170 ---SVQLELDQSGTW---TCTVLQ--NOKVFKIDV---VCPAPRPSCKDKHTCP 217
 Db 199 SGFYSLSLVTPSSSTWPSQGYVCNVAHPASKTELKRIEPRIPKSTPPGS-----SCP 253
 Oy 218 --ELLGSPVFLPEPPKCDTLMISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVHNAKTK 275
 Db 254 PGNILGGSPVFLPEPPKCDTLMISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVHNAKTK 313
 Oy 276 PREEQVNSTYRVVSVLTVLHOMLNGKEYKCKVSNKALPAPIEKTIISKAKGQPREPOVY 335
 Db 314 PREAQNSTYRVVSVLTVLHOMLNGKEYKCKVSNKALPAPIEKTIISKAKGQPREPOVY 373
 Oy 336 LPPSPDELTKNQVSLTCLVKGFYPSDIAVEMESNGOPENNYKTTPPVLDSDGSFFLYS 395
 Db 374 IEPPEKQMKKKYSLTCLVTFNPFSSEAIIVDEWENGLIEDQYKATPILDSGTFILYSK 433
 Oy 396 TVDKSRMQGVNFCSCVMHEALHNHYTOKSLSPG 431
 Db 434 TVDTDSWLGGEIFTCGVVHEALHNHYTOKSLSPG 469
 RESULT 11
 O9R1A4 PRELIMINARY; PRT; 437 AA.
 ID O9R1A4;
 AC O9R1A4;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Gamma1 heavy chain of Mab7 (fragment).
 GN IGH-4.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilde K.G., Yu X., Ekrumodoullah A.K.M., Misra S.;
 RT "Cloning of cDNAs encoding for anti-white pine blister rust monoclonal
 RT antibody (Mab 7, its light and heavy chains) and construction of a
 RT single chain antibody (scFv).";
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF152372; AAD40243.1; -
 DR PIR; B45837; B45837.
 DR PDB; 1COK; 11-SEP-99.
 DR PDB; 1191; 25-DEC-02.
 DR PDB; 1KCU; 11-MAY-02.
 DR MGD; MGI:96446; Igh-4.
 DR InterPro; IPR007110; Igh-1like.
 DR InterPro; IPR003006; Igh_MHC.
 DR InterPro; IPR003596; Igh_V.
 DR Pfam; PR00047; Igh_3.
 DR SMART; SM00406; Igh_1.
 DR PROSITE; PS00835; IGH_LIKE; 4.
 DR PROSITE; PS00290; IGH_MHC; 1.
 FT NON TER 1
 FT NON TER 437
 SQ SEQUENCE 437 AA; 48142 MW; 5C3A7BB3E7D697C CRC64;
 Query Match 35.9%; Score 867.5; DB 11; Length 437;
 Best Local Similarity 41.6%; Pred. No. 3.5e-64;
 Matches 188; Conservative 66; Mismatches 123; Indels 75; Gaps 11;
 Oy 30 LGKGGDTVELTCTAS--QKSIQFHWKNSNQIKILNGOSF-LTKGPKSLNDPADRSRL 86
 Db 10 LKPGASVATSCAASGYTGTGYMHWVSKSHGSLWGLVNPBNDGTITNKGFKKATL 89
 Oy 88 ---DGNFPLIKNIKIEDSDTYICEVEDQKEVOLVFGLTANSDTHLQGSLLTL 144
 Db 70 KDKDNILSLQMSLSRSEDITANYC-----AKRDSAYMGPGLVTVS 112
 Oy 145 SPGSSPSVQCRSP-----RKNIGGKTL--SVSQL 174
 Db 113 AKTTPPSVPLAPSAQTNMVTGLGVKGYFPEPYTVKNNGALSSGVATYSSVLQ 172
 Oy 175 ELQ-----DSGTW---TCTVLQ--NOKVFKIDV---VCPAPRPSCKDKHTCP 219
 Db 173 DLYTSSSVTPSSSTWPSQGYVCNVAHPASKTELKRIEPRIPKSTPPGS-----SCP 227
 Oy 220 LGGPSPVFLPEPPKCDTLMISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVHNAKTKPRE 279
 Db 228 ---SVFIFPPKPKVLTITLTLPKVTCTVVDVSKDPEVQFSMFPVDDVEVHTAQTQPRE 284
 Oy 280 QVNSTYRVVSVLTVLHOMLNGKEYKCKVSNKALPAPIEKTIISKAKGQPREPOVY 339
 Db 286 QVNSTYRVVSVLTVLHOMLNGKEYKCKVSNKALPAPIEKTIISKAKGQPREPOVY 344
 Oy 340 RDELTKNQVSLTCLVKGFYPSDIAVEMESNGOPENNYKTTPPVLDSDGSFFLYS 399
 Db 345 KEQMKDKRSLTCLMTDFPEPDITVEQMONGPAPNRYKNTQIPMDTDSYFYASKLVNQ 404
 Oy 400 SRMQGVNFCSCVMHEALHNHYTOKSLSPG 431
 Db 405 SNWEAGNTFTCSVLHEGLHNHYTOKSLSPG 436
 RESULT 12
 O8R3V9 PRELIMINARY; PRT; 469 AA.
 ID O8R3V9;
 AC O8R3V9;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN IGH-4.

OS Mus musculus (Mouse).
 OC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_TaxId=10090;
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC024405; AAH24405.1; -
 DR PIR; B45837; B45837
 DR MGI; MGI:96446; Igh-4.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig_3.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 469 AA; 51976 MW; 534793F155D05457 CRC64;

Query Match 35.6%; Score 858.5; DB 11; Length 469;
 Best Local Similarity 41.6%; Pred. No. 2.2e-63;
 Matches 192; Conservative 67; Mismatches 120; Indels 83; Gaps 14;

```

QY 30 LGKGGDTVELTCTAS-----QKKSIOFHMKNSNQIKILG--NOGSFLTKGPS 74
DB 30 LVPGSGSLRLSCAAGFTFTDYMSWVRPQKALEWLGFIKANKANGTTESSAVKRF 89
QY 75 KLNRADSRRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVGLTANS DTHLL 134
DB 90 TIS--RDNSQSILYLQMN-----ALRAEDSATYYC--ADRRSSYYY-----SGTSFAY 134
QY 135 OGOSLTLTLESPPGSSPSVQCNSP-----RGNKIQG 165
DB 135 WGGGTLVTVSAKKTTPSVYPLAPGSAAGTNSMTVLGCLVKGYFPEPVTVTNSSLSG 194
QY 166 GKTL-SVSOLELO-----DSGTW-----TCTVLQ--NOKVFEKIDIVPCAPAPRKS 209
DB 195 VHTFPAVLQSDLYTLSSSVTVPSSTWPGQTVTCNVAHAASTKVDKIVPRDCCG-KPCI 253
QY 210 CDKHTTCPELLGSPVFLFPKPKDTLMSRTPEVTCVAVDVSHEDPEVKENMYDGEV 269
DB 254 C---TVPEV---SSVFIFPPKPKDVLITTLTPKVTGVVDISKDPEQFSWFEDVDEV 306
QY 270 HNAKTRPEEOVNSTYRVVSVLTIVLHOMLNGEKYCKVSNKALPAPIEKTIKAKGQPR 329
DB 307 HTAQTRPEEOVNSTYRVVSELPIWHQDLNGEKYCKVSNKALPAPIEKTIKAKGQPR 366
QY 330 EPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAYEWESNGQPENNYKTTTPVLDSGSP 389
DB 367 APQVYTLPPKQMAKDKVSLTCTMTDFPEDITTEMQNNQPAENYKNTQPIMDTGSY 426
QY 390 FLYSKLTVDKSRMQGNVFGSCVMEHALNHYTQKSLSPG 431
DB 427 FVYSKLTAVQKSNMEAGNTFTCSVLHEGLHNHHTKSLSPG 468

```

RESULT 13

Q99LC4 PRELIMINARY; PRT; 463 AA.
 AC 099LC4;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Similar to RIKEN CDNA 1810060009 gene.
 GN IGH-4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_TaxId=10090;
 RP SEQUENCE FROM N.A.

RA Strauberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC003435; AAH03435.1; -
 DR PIR; B45837; B45837.
 DR HSSP; P01842; 7PAB.
 DR MGI; MGI:96446; Igh-4.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig_3.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 SQ SEQUENCE 463 AA; 51007 MW; EAA674C6B8C30783 CRC64;

Query Match 35.1%; Score 848.5; DB 11; Length 463;
 Best Local Similarity 40.6%; Pred. No. 1.5e-62;
 Matches 191; Conservative 63; Mismatches 111; Indels 105; Gaps 15;

```

QY 30 LGKGGDTVELTCTAS-----QKKSIOFHMKNSNQIKILG--SPLTKG 72
DB 30 LARGCAVRLSCAKSGYFTFGYVSWKQRTQGLEWVG---EIVGSGNTYYSEKFKG 85
QY 73 PSKI--NDRADSRRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVGLTANS-- 129
DB 86 KATLTTRKSSSTAYMH-----LSLTSEDSAVYFC-----ARSSY 120
QY 130 ---DTHLLQOSLTLTLESPPGSSPSVQCNSP----- 158
DB 121 YSYDLFAVWGGTILVTVSAKKTTPSVYPLAPGSAAGTNSMTVLGCLVKGYFPEPVTVTW 180
QY 159 -RGNKIQGKTL-SVSOLELO-----DSGTW-----TCTVLQ--NOKVFEKIDIVP 201
DB 181 NSSGSLSSGVHTFPAVLQSDLYTLSSSVTVPSSTWPGQTVTCNVAHAASTKVDKIVPRD 240
QY 202 CPAPEPKSCDKHTTCPELLGSPVFLFPKPKDTLMSRTPEVTCVAVDVSHEDPEVKEN 261
DB 241 CGC-KPCIC---TVPEV---SSVFIFPPKPKDVLITTLTPKVTGVVDISKDPEVOFS 292
QY 262 MYVGVGVHNAKTRPEEOVNSTYRVVSVLTIVLHOMLNGEKYCKVSNKALPAPIEKTI 321
DB 293 WFDVDEVHTAQTRPEEOVNSTYRVVSELPIWHQDLNGEKYCKVSNKALPAPIEKTI 352
QY 322 SKAGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAYEWESNGQPENNYKTTTP 381
DB 353 SKTGGRKAPQVYTLPPKQMAKDKVSLTCTMTDFPEDITTEMQNNQPAENYKNTQPI 412
QY 382 VLDSGSPFLYSKLTVDKSRMQGNVFGSCVMEHALNHYTQKSLSPG 431
DB 413 IMDTGSYFVYSKLTAVQKSNMEAGNTFTCSVLHEGLHNHHTKSLSPG 462

```

RESULT 14

Q99LJ1 PRELIMINARY; PRT; 468 AA.
 AC 099LJ1;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Similar to RIKEN CDNA 1810060009 gene.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_TaxId=10090;
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC003878; AAH03878.1; -
 DR PDB; 2AP2; 24-NOV-99.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.

DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 SQ SEQUENCE 468 AA; 51661 MW; 96352328B332ADB CRC64;

Query Match 35.0%; Score 846; DB 11; Length 468;
 Best Local Similarity 42.7%; Pred. No. 2,4e-62;
 Matches 198; Conservative 55; Mismatches 123; Indels 88; Gaps 15;

QY 30 LGKGDVTELTCTAS--QKKSIOFHMKNSNOIKILNQG-----SFLTKGPSKLN 77
 DB 30 LVKPGASVKLSCTASGFNIKIDSLMHV-----VKOREQGLIEWIGWIDPDEGKVPKPG 84
 QY 78 DDAADSRSLMDQGNFLLIKNLKIEDSDTYICEVEQKEVQLVFGLTANSPTHLQOQ 137
 DB 85 DKATTTADTSSNTAY-LQLSLSTSEDTALYYC-----ARNLVAG-----GYDYWQ 130
 QY 138 STLTLSPSPGSSPSVQ-----CRSPRGKNIQGG-----KTL----- 169
 DB 131 GTTITVSSAKTTAPSYPLAPVCGDTTSSVTLGCLVKGYPPEVTLTNMSSLSGVT 190
 QY 170 --SVSQLEIQ-----DSGTW-----TCTVLQ--NOKKVEFKID-----IVPCPAPP 207
 DB 191 PFAVLQSDLYTLSSSVTVTSSSTWPSQSLTCNVNHPASSSTKVDKIDPRGPITKPCP--P 247
 QY 208 KSCDKTHTPELLGGSVFLFPKPKDITLMISTPEVTCVVDVSHEDPEVKFNMYVDGY 267
 DB 248 CRC---PAPNLLGGSVFIFPKKIDVLMISLSPWTCVVDVSDDDPDVQISWFNVNV 303
 QY 268 EVHNAKTPREBOYNSTYRVSVLTVLHODMLNGKCKVSKVSKALPAPIEKTISKAKQ 327
 DB 304 EULTAQTQTHREDYNSTLRAVVSALPIQHDQMSGKFKCKVNNKALPAIERTISKPKS 363
 QY 328 PREQVYTLPPSRDELTKNQVSLTCLVKGYPSPDIAVEMWESNGQEPNNYKTPPVLDSDG 387
 DB 364 VBAQVYVLPPEBEWTKQVTLTCVTDPFMEDIVVEMTNNKGTBLANTKEPVLDSG 423
 QY 388 SFPLYSKLTVDKSRWQGNVFGSVHAEALHNHYTKSLSPG 431
 DB 424 SYFMYSKLVKVEKNWERNSSCSVHAEGLHNHHTTKSPSRTPG 467

RESULT 15

Q9DBL4 PRELIMINARY; PRT; 473 AA.
 AC Q9DBL4;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE 181006009RIK protein.
 GN IGH-1 OR 181006009RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Pancreas;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kondo S., Yamahara T.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann M., Gaasterland T., Gliesi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Stadl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guetlich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
 RA Lyons P., Marchionni L., Maehima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuke S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK007918; BAB25349.1; --
 DR PIR; S26746; S26746.
 DR HSSP; P01842; 7FAB.
 DR MGD; MG1:96443; Igh-1.
 DR InterPro; IPR007110; Igh-1-like.
 DR InterPro; IPR003006; Igh_MHC.
 DR InterPro; IPR003596; Igh_v.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 SQ SEQUENCE 473 AA; 51699 MW; 9DED57A514475FBB CRC64;

Query Match 34.5%; Score 833.5; DB 11; Length 473;
 Best Local Similarity 40.0%; Pred. No. 2.8e-61;
 Matches 185; Conservative 67; Mismatches 131; Indels 79; Gaps 11;

QY 30 LGKGDVTELTCTASQKKSIOFHMKNSNOIKILNQG-----SFLTK-----GPSKLNDRADSR 83
 DB 30 LVKPGASVKLSCTASGTYFTDY--IMVKOREQGLIEWIGKIPGSGSTYNEKFKCK 86
 QY 84 RSLW--DQGNFLLIKNLKIEDSDTYICEVEQKEVQLVFGLTANSPTHLQOQSILT 140
 DB 87 ATLTKDSSSTAYWGLSLTSEDSAVYFCASGY-----DYDWFAYWQOGLT 133
 QY 141 LTLSPSPGSSPSVQ-----CRSPRGKNI----- 163
 DB 134 VIVSAKTTAPSYPLAPVCGDTTSSVTLGCLVKGYPPEVTLTNMSSLSGVTPEPA 193
 QY 164 ---OGKTLVSQLELDQSGTW-----TCTV-----LONOKVEFKIDIVPCPAPPKS 209
 DB 194 LIQSLYTLSS--VTTSNTWPSQSLTCNVNHPASSSTKVDKIDPRPITONPCPPLKE 251
 QY 210 CDKTHTECELLGGSVFLFPKPKDITLMISTPEVTCVVDVSHEDPEVKFNMYVDGYEV 269
 DB 252 CPKC--AAPDLGGPSVFIFPKKIDVLMISLSPWTCVVDVSDDDPDVQISWFNVEV 310
 QY 270 HNAKTPREBOYNSTYRVSVLTVLHODMLNGKCKVSKVSKALPAPIEKTISKAKQPR 329
 DB 311 HTAQQTQTHREDYNSTLRAVVSALPIQHDQMSGKFKCKVNNKALPAIERTISKPRGPR 370
 QY 330 EPOVYTLPPSRDELTKNQVSLTCLVKGYPSPDIAVEMWESNGQEPNNYKTPPVLDSDSF 389
 DB 371 APOVYVLPPEBEWTKKESLTCMTGFLPAEIAVDMWNSNGTEQNYKNTATVLDSDSY 430
 QY 390 FLYSKLTVDKSRWQGNVFGSVHAEALHNHYTKSLSPG 431
 DB 431 FMYSLTRQKSTWERSGLFACSVHAEGLHNHHTTKTISRSLG 472

RESULT 16

Q99L25 PRELIMINARY; PRT; 473 AA.
 AC Q99L25;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to RIKEN CDNA 181006009 gene.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.


```

Db      61 DOGCSMTIIKNLKIDSEDTYICEVENKKEEVLVFGILTANSDDLHLOGOSLTLTLESPP 120
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      148 GSSPSVOCRSPRGKNIQGGKTLVSQLELQDSGTWCTVYLNQKQKVEPRIDIV 200
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 GSSPSVCKRSPRGKNIQGGRTLSVPOLERQDSGTWCTVSDQNTVEFKIDIV 173
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 21

```

009263 ID 009263 PRELIMINARY: PRT: 397 AA.

```

```

AC 009263; 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DR 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxId=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Machiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001222; AAB60869.1; -.
DR HSSP; P01730; 1W1Q.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43994 MW; A3CD031535A51524 CRC64;

```

```

Query Match      31.8%; Score 767; DB 6; Length 397;
Best Local Similarity 85.5%; Pred. No. 8.1e-56;
Matches 148; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

```

```

Qy      28 VVLGKKGDTVELTCTASQKKSIOFHKNSNOIKILGNQGSFLTKGPSKLNDRADSRSLW 87
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 VVLGKKGDTVELTCTASQKKSIOFHKNSNOIKILGNQGSFLTKGPSKLNDRADSRSLW 60
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      88 DOGNPFLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDDLHLOGOSLTLTLESPP 147
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 DOGCSMTIIKNLKIDSEDTYICEVENKKEEVLVFGILTANSDDLHLOGOSLTLTLESPP 120
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      148 GSSPSVOCRSPRGKNIQGGKTLVSQLELQDSGTWCTVYLNQKQKVEPRIDIV 200
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 GSSPSVCKRSPRGKNIQGGRTLSVPOLERQDSGTWCTVSDQNTVEFKIDIV 173
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 22

```

009263 ID 009263 PRELIMINARY: PRT: 397 AA.

```

```

AC 009263; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DR 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4
OS Cercopithecus pygerythrus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxId=60710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Machiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001222; AAB60874.1; -.
DR HSSP; P01730; 1W1Q.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43994 MW; 21C3E30882ABFBC0 CRC64;

```

```

Query Match      31.8%; Score 767; DB 6; Length 397;
Best Local Similarity 85.5%; Pred. No. 8.1e-56;
Matches 148; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

```

```

Qy      28 VVLGKKGDTVELTCTASQKKSIOFHKNSNOIKILGNQGSFLTKGPSKLNDRADSRSLW 87
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 VVLGKKGDTVELTCTASQKKSIOFHKNSNOIKILGNQGSFLTKGPSKLNDRADSRSLW 60
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      88 DOGNPFLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDDLHLOGOSLTLTLESPP 147
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 DOGCSMTIIKNLKIDSEDTYICEVENKKEEVLVFGILTANSDDLHLOGOSLTLTLESPP 120
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      148 GSSPSVOCRSPRGKNIQGGKTLVSQLELQDSGTWCTVYLNQKQKVEPRIDIV 200
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 GSSPSVCKRSPRGKNIQGGRTLSVPOLERQDSGTWCTVSDQNTVEFKIDIV 173
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 23

```

009259 ID 009259 PRELIMINARY: PRT: 397 AA.

```

```

AC 009259; 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DR 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4
OS Cercopithecus sabaues.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxId=60711;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Machiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001223; AAB60870.1; -.
DR HSSP; P01730; 1W1Q.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.

```



```

SQ SEQUENCE 397 AA; 43980 MW; F74C42E22B196155 CRC64;
Query Match 31.4%; Score 758; DB 6; Length 397;
Best Local Similarity 85.0%; Pred. No. 4.6e-55;
Matches 147; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 28 VVLGKGGDTVELTCTASQKKSIOFHMKNSNQIKILGNQGSFLTKGPSKLNDRADSRSLW 87
DB 1 VVLGKGGDTVELTCTASQKTTQTFHMKNSNQIKILGNQGSFLTKGPSKLNDRIDSRSLW 60

QY 88 DQGNPFLIIKNIKIDSDPTIYCEVEDQKEVOLLVFGLTANSPTHTLLOGSITLTLESPP 147
DB 61 DQGSCEMIILNIKIDSETTYICEVENKKEVELLVFGLTANSPTHTLLOGSITLTLESPP 120

QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTCTCTVLOKQKVEFKIDIV 200
DB 121 GSSPSVKCRSPRGKNIQVGRITLSVPLERQDSGTCTCTVSDQNTVERKIDIV 173

RESULT 26
Q8H2T8 PRELIMINARY; PRT; 457 AA.
AC Q8H2T8;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Lymphocyte antigen CD4.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrich.
OC NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22174698; PubMed=12186836;
RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;
RT "Blockade of HIV-1 Infection of New World Monkey Cells Occurs
RT Primarily at the Stage of Virus Entry.";
RL J. Exp. Med. 196;431-445(2002).
DR EMBL; AF452616; AAN14532.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00409; IG_3.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
SQ SEQUENCE 457 AA; 50878 MW; 718CFDB78D97F59B CRC64;

Query Match 30.8%; Score 743; DB 6; Length 457;
Best Local Similarity 46.7%; Pred. No. 1e-53;
Matches 178; Conservative 45; Mismatches 96; Indels 62; Gaps 10;

QY 1 MNRGVPRHLLLVQLALPPAATQGNKVLGKKGDTVELTCTASQKKSIOFHMKNSNQIK 60
DB 1 MNGGIPFRHLLLVQLALLIAVTHGKTIVLGKGEVVELPCTSLKXKLQPHMKTSNQIK 60

QY 61 ILGNQGSFLTKGPSKLNDRADSRSLWDQGNFPIIINKIKIEDSDTYICEVEDQKEVQL 120
DB 61 ILGIGQSFYTKGQSKLNRIDSKSSWDRGSPFLIIRNVQVEDSETTYICEVSKKEVEL 120

QY 121 LVFGLTANSPTHTLLOGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 QVFGLTANPDHTLLOGSITLTLESPPGSSPSVECSPPRGKTIKRMKTLFWSQLTIQDSG 180

QY 181 TWTCTVLOKQKVEFKIDIVPCPAPEPKSCDKTHICPELLGSPVFLFP----- 229
DB 181 TMKCTVSOHLELV-FKINIVVLAFOQAS-----IVYKKEGQGVFSPFLAFAQLGSG 234
QY 230 -----PKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMVYDVGVEVHNAKTKPRE- 278
```

```

DB 235 GELCMAERASSSSSWITFLTNQEV--VVLVYQDPKLRM-----CKKLPRHLLTLPQAL 287
QY 279 EGY-----NSTRVVSVLTVLHODWLNGKEYK-----CVY----SNKALPAPI 317
DB 288 PÖYAGSGFTLALGKTGKHQ-----EVLVVRATPOLONNLTCEVWGFTSPGLVLSL 341

QY 318 EKTISKAKGQPREPOVYTLPP 338
DB 342 KLENQEAQSKREKAVVWLVNP 362

RESULT 27
Q8H2T7 PRELIMINARY; PRT; 457 AA.
AC Q8H2T7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Lymphocyte antigen CD4.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OC NCBI_TaxID=9521;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22174698; PubMed=12186836;
RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;
RT "Blockade of HIV-1 Infection of New World Monkey Cells Occurs
RT Primarily at the Stage of Virus Entry.";
RL J. Exp. Med. 196;431-445(2002).
DR EMBL; AF452617; AAN14533.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00409; IG_3.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
SQ SEQUENCE 457 AA; 50899 MW; B164DA89E70C575A CRC64;

Query Match 30.7%; Score 740.5; DB 6; Length 457;
Best Local Similarity 72.0%; Pred. No. 1.6e-53;
Matches 144; Conservative 24; Mismatches 31; Indels 1; Gaps 1;

QY 1 MNRGVPRHLLLVQLALPPAATQGNKVLGKKGDTVELTCTASQKKSIOFHMKNSNQIK 60
DB 1 MNGGIPFRHLLLVQLALPPATHGKTIVLGKGEVVELPCTSLKXKGVPHMKTSNRK 60

QY 61 ILGNQGSFLTKGPSKLNDRADSRSLWDQGNFPIIINKIKIEDSDTYICEVEDQKEVQL 120
DB 61 ILGVQNVFVTRGQSKLNRIDSKSSWDRGSPFLIIDARIBSETTYICEVSKKEVEL 120

QY 121 LVFGLTANSPTHTLLOGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 QVFGLTANPDHTLLOGSITLTLESPPGSSPSVECTSPRGKIRIGKRTLSVSQIGIPDSG 180

QY 181 TWTCTVLOKQKVEFKIDIV 200
DB 181 TMKCTVFOHLELV-FEINIV 199

RESULT 28
Q9XS78 PRELIMINARY; PRT; 455 AA.
AC Q9XS78;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
```

DE T-cell surface glycoprotein CD4.
 OS Delphinapterus leucas (Beluga whale).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
 OC Monodonta; Delphinapterus.
 OX NCBI_Taxid=9749;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymus;
 RX MEDLINE=99216435; PubMed=10199913;
 RA Romano T.A., Ridgway S.H., Felten D.L., Quaranta V.;
 RT "Molecular cloning and characterization of CD4 in an aquatic mammal,
 the white whale Delphinapterus leucas.";
 RL Immunogenetics 49:376-383(1999).
 DR EMBL: AF071999; AAD23738.1; -.
 DR HSSP: P01730; 1W1Q.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0006955; P:immune response; IEA.
 DR InterPro: IPR000973; CD4_TcAg.
 DR InterPro: IPR007110; Ig_Like.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig_3.
 DR PRINTS: PR00692; CD4TCANTIGEN.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PS50835; IG_Like; 2.
 SQ SEQUENCE 455 AA; 50499 MW; AA532FD4411AASD1 CRC64;

Query Match 24.9%; Score 601.5; DB 6; Length 455;
 Best Local Similarity 40.8%; Pred. No. 7.1e-42;
 Matches 155; Conservative 43; Mismatches 107; Indels 75; Gaps 9;

QY 1 MNRGVPFRHLVLVLTALLPAATQGNKVLGKGGDVVELTCTASQKSIQPHMKNNOIK 60
 DB 1 MDRSTSLRHLFLVLQVLMPLAGTQGNKVLGKAGLAELEPCASQNKSLFFSKNSYQTK 60
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIITKNLKIENSDTYICEVEDQKEEYOL 120
 DB 61 ILGRHYFVHMKGSNLHSRVESKINLMDQSPFLVKDLLEVPDSGYICEVEDEKKEIEVL 120
 QY 121 LVFGLTANSDTHLLQGSILTLTIESPPGSSPSVQCSPPRKNIQGGKTLVSQLELDGSG 180
 DB 121 QVFRLLTASSSDTRLRLQLQGSILTLTLEGPSSGNPSVQMGKPGNKRNKAKSLPLQVGLQDGG 180
 QY 181 TMTCTVLOKOKKVEFKIDIVPCAPAPKSCDKHTCEPELLGGSVLPFPKPKDTLMISR 240
 DB 181 TMTCTVSOAQQTIVFNKHLVLAFOEVSS---TVYAKEGEQMNFSPF----- 224
 QY 241 TREVTCVVDVSHDEPEVKNFNVYDGVNHNNAKTKPREQVNSTYRVSVLFTVLHODWL- 299
 DB 225 -----LTFGDENLSGELSWL-----QAQKNS-----PESWIT 252
 QY 300 ---NG-----KEYCKVSNKALPAPIEKTIISAKQOPREPOVYTLPPSRDEL--- 343
 DB 253 FKLNNKVTYVGRKARXKDLKLRMS-KAL-PLHLTLPLQALPQVAGSGNLTLNLTGKLYQEV 309
 QY 344 -----TKQVSLTCLVKG 356
 DB 310 NLVVMKVTSPNSLTCEVLG 329

RESULT 29
 P79355 PRELIMINARY; PRT; 474 AA.
 AC P79355;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CD4 antigen precursor.
 OS Felis silvestris catus (cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_Taxid=9685;
 RN (1)

RP SEQUENCE FROM N.A.
 RA Miyazawa T.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92165316; PubMed=1537604;
 RA Norimine J., Miyazawa T., Kawaguchi Y., Tohya Y., Kai C., Mikami T.;
 RT "A cDNA encoding feline CD4 has a unique repeat sequence downstream of
 the V-like region.";
 RL Immunology 75:74-79(1992).
 DR EMBL: AB000483; BAA19124.1; -.
 DR HSSP: P01730; 1W1Q.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0006955; P:immune response; IEA.
 DR InterPro: IPR000973; CD4_TcAg.
 DR InterPro: IPR007110; Ig_Like.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig_3.
 DR PRINTS: PR00692; CD4TCANTIGEN.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PS50835; IG_Like; 2.
 DR SIGNAL.
 FT CHAIN 1 26 POTENTIAL.
 FT SIGNAL. 27 474 POTENTIAL.
 SQ SEQUENCE 474 AA; 52243 MW; D946DD4BEAD00EC CRC64;

Query Match 23.1%; Score 557.5; DB 6; Length 474;
 Best Local Similarity 55.3%; Pred. No. 3.7e-38;
 Matches 121; Conservative 30; Mismatches 49; Indels 19; Gaps 3;

QY 1 MNRGVPFRHLVLVLTALLPAATQGNKVLGKGGDVVELTCTASQKSIQPHMKNNOIK 59
 DB 1 MNOGAVFRHLVLVLTALLPAATQGNKVLGKAGLAELEPCASQNKSLFFSKNSYQTK 60
 QY 60 KILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIITKNLKIENSDTYICEVEDQKEEYOL 118
 DB 61 KILGSHSLCLTGSKLTRESKILMDQSPFLVKDLLEVPDSGYICEVEDEKKEIEVL 120
 QY 119 QLVFGLTANSD-----THLLQGSILTLTIESPPGSSPSVQCSPPRKX 161
 DB 121 ELTVFGLTAKVDPSSGSGSSSTSTSTSYLLQGSILTLTIESPSSNPVSQMGKPGNK 180
 QY 162 NIQGGKTLVSQLELDGSGTGTCTVLOKOKKVEFKIDIV 200
 DB 181 SKSGVSHSLSLQLEBESGTCCTIVSQSQKTLVFNITL 219

RESULT 30
 O61396 PRELIMINARY; PRT; 457 AA.
 AC O61396;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE T-cell differentiation antigen.
 GN CD4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8041159; PubMed=2823269;
 RA Gorman S.D., Tourville B., Parnes J.R.;
 RT "Structure of the mouse gene encoding CD4 and an unusual transcript in
 brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
 DR EMBL: M17080; AAA37402.1; JOINED.
 DR EMBL: M17076; AAA37402.1; JOINED.
 DR EMBL: M17077; AAA37402.1; JOINED.
 DR EMBL: M17078; AAA37402.1; JOINED.
 DR EMBL: M17079; AAA37402.1; JOINED.
 DR HSSP: P01730; 1W1R.

Matches 75; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 107 YICEVEDKEEYVQLVFGLTANSPTHLLOGSLTLTLPSSSPSSVQCRSPRGKNIQGG 166
DB 1 YICEVEDKEEYVQLVFGLTANSPTHLLOGSLTLTLPSSSPSSVQCRSPRGKNIQGG 60

QY 167 KTLVSQLELQDSGTWTCTVLQNGK 191
DB 61 RTLSPQLERQDSGTWTCTVSDQK 85

RESULT 36
Q96EY0
ID Q96EY0 PRELIMINARY; PRT; 613 AA.

AC 096EY0:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_taxid=9606;
OK NCBI_taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=cell;
RL Struhsberg R;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011857; AAH11857.1; -.
DR PIR; S15590; S15590.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS50835; IG_LIKE; 5.
DR PROSITE; PS50290; IG_MHC; 3.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 613 AA; 67273 MW; 31214203FB8421E7 CRC64;

Query Match 16.4%; Score 395.5; DB 4; Length 613;
Best Local Similarity 22.2%; Pred. No. 1.9e-24;
Matches 135; Conservative 83; Mismatches 206; Indels 185; Gaps 19;

QY 8 RHLLVLOLALP-----AATQKQKVLGKKGDVLELTCTASQKKSIOFHMKNSNOIKI 61
DB 2 KHLWFFLLVVAAPRVNLQVQLQESGPGLVKPSFTLSLTCTVSGSISSTYMSWIRQPAQ 61

QY 62 LGNQ--GSFLTQKPSKLNDRADSRSL--WDQGFPLIKNLKIEDSDTYIC----- 109
DB 62 KGLEWIGRIYTGSTINYNPSLKSRYVTMSVDTSKQPSLKSVTAAADTAIVYCAQPMEL 121

QY 110 -----EVEDQKEEYVQLVFGLT----- 127
DB 122 PTVGFLFWGQGLTVTVSSGSASAPTLFPLVSCNSPSPDSSVAVAGLAQDFLPDSITPSW 181

QY 128 ----NSD-----THLQGS----- 138
DB 162 KYKNSDLSISTQFSSVLRGKRYATSOVLPSKVMQGTDEHVCKVQHPNGKKNV 241

QY 139 LTLTLPSSPGSSPSVQCR-----SPR-----GKNIQGGKTL 170
DB 242 LPVIALPPLKVSVPFPRDGFPGFNPRKSKLIGQATGFSRQIOWMLBKGQVSGVTTD 301

QY 171 VSQLELQDSG-----TWCTVLQNGKVEFKIDYPCAPPEKS 209
DB 302 QVQAAKESGPTTYKVTSTLTIKESDMLSQSMFTGRV--DHRGLTFQONASSMCPDDDT 359

QY 210 CDKHTCEPLLQSGSPVFLPPPKKOTLMSRTPEVTCVVDVSHEDPEYKFNWYDVGEV 269
DB 360 AIR-----VFALPSP-FASIFLTSTKTLCTLVTDLTYYD-SVTTISWTRQNGEA 405

QY 270 HNAKTPREEOYNSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAKGP- 328

DB 406 VKHTNTNSESHPNATFPAVAGASICEDDMNSGEFTCTVHTDLPSPKQTLISRPKVAL 465

QY 329 REPQVYLLPSPRDEL-T-KNOVSLTCLYKGPSPSIAMWESNGDP--ENNYKTPPVLD- 384
DB 466 HRPVYLLPSPAREQINRESATITCLVTGSPADVPQOMRGQPLSPKRYTAPPEP 525

QY 385 -SDSPFLYSKLTVDKSRMQGNVFCSSVMHEALHNHYTOKSLSPQLQDFTCAKQD 443
DB 526 QAPGRYFAHSILITSESEKMTGETYTCVVAHEALPNRYTERTVKS-----TEGEVSADE 580

QY 444 GELDGLMTT 452
DB 581 EGFENLMAT 589

RESULT 37
O77595
ID O77595 PRELIMINARY; PRT; 86 AA.

AC O77595:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Cercopithecus galerellus chrysogaster.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Cercopithecidae;
OK NCBI_taxid=75569;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; Pubmed=9656488;
RA Harris E.E., Dieotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (primates: Papionini)." ;
MoJ. Biol. Evol. 15:892-900 (1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DB EMBL; AF057382; AAC5126.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR007110; IG_1like.
KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON TER 1
FT DOMAIN 1
FT 1 19 IG-LIKE V-TYPE DOMAIN.
FT 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CANBOHD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFD 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9419 MW; A9D97A9EB19582BE CRC64;

Query Match 16.4%; Score 395; DB 6; Length 86;
Best Local Similarity 88.2%; Pred. No. 1.4e-23;
Matches 75; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDKEEYVQLVFGLTANSPTHLLOGSLTLTLPSSSPSSVQCRSPRGKNIQGG 166
DB 1 YICEVEDKEEYVQLVFGLTANSPTHLLOGSLTLTLPSSSPSSVQCRSPRGKNIQGG 60

QY 167 KTLVSQLELQDSGTWTCTVLQNGK 191
DB 61 RTLSPQLERQDSGTWTCTVSDQK 85

RESULT 38
O77598
ID O77598 PRELIMINARY; PRT; 86 AA.

DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_V.
 DR Pfam; PF00047; IG; 5.
 DR SMART; SM00406; IG; 1.
 DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
 DR PROSITE; PS00835; IG_LIKE; 5.
 DR PROSITE; PS00290; IG_MHC; 3.
 DR Hypothetical protein.
 KM Hypothetical protein.
 SQ SEQUENCE 614 AA; 67921 MW; 55EF536E77AA9BBB CRC64;

Query Match 16.1%; Score 389.5; DB 4; Length 614;
 Best Local Similarity 23.9%; Pred. No. 6.2e-24;
 Matches 116; Conservative 77; Mismatches 167; Indels 125; Gaps 16;

```

QY 36 TVELCTTASQ--KKSIOFHMKNSNOIKILGNQGSFLTGPSKLNDRADSRSSIMDQGNFP 93
DB 163 SVAAGCCLADFLPDSITFSWKYKNNSDISSTRG-----FP 197
QY 94 LIINKLKIEDSDTYICEVEDQKEVQLLVGLTANSPTLL-----QGQSILTT 142
DB 198 SVLRGCKXAATSOVLPSKD-----VMQGTDEHVCKVQHNGKKNVPLPI 246
QY 143 LESPSSPSVQCR-----SPR-----GKNIQGGKTLVSQ 174
DB 247 AELPKVSVFVPPRDGFNGNPKRSKLICQATGSPROIQVSWLRGKQVSGVTTDQVQA 306
QY 175 ELQDSG-----TWCTVLAQNKVKEFKIDIVPCAPAPKSCDKT 213
DB 307 EAKESGPTTYKKTSTLTIKESDMLSGMFTCRV--DHRGLTFQONASMCVDDQTAIR- 363
QY 214 HTCELLGGSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFMVVDGVEVHNAAK 273
DB 364 -----VFALPPS-PASIFLTSTKTLCLVLTDLTTYD-SVLTISWRONGEAVKTH 410
QY 274 TKPREQYNSTYRVSVLTVLHODWLNKKEYCKKSNKALPAPIKTISSAKQGP-REPO 332
DB 411 TNISSEHNATFSAVGSASICEDDMNSGSRFTCTVTHDLPSPFKQIISRKGVALHRPD 470
QY 333 VYTLPPSRDEL-T-KNOVSLTCLVKGFPYPSDIAVWESNGQP--ENNYKTTPEVLD--SDG 387
DB 471 VYLLPPAREQLNLRSAITITCLVTGFSPADVFVQMORGQPLSPKQVTSAPMEPEQAPG 530
QY 388 SFELYSKLTVDKSRMOQGNVSCSVMEHALHNHYTQKSLSLSPGLQDCECAEQDDEL 447
DB 531 RYFAHSILTVSEENGTGETYTCVAHEALPNRYETRTVDS-----TEGEVSADDEGFE 585
QY 448 GLWTT 452
DB 586 NLMAT 590

```

RESULT 41

Q96AA6 PRELIMINARY; PRT; 618 AA.
 AC Q96AA6;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strauberg R.;
 RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC017356; AAH17356.1; -
 DR PIR; S15590; S15590.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_V.

DR Pfam; PF00047; IG; 5.
 DR SMART; SM00406; IG; 1.
 DR PROSITE; PS00835; IG_LIKE; 5.
 DR PROSITE; PS00290; IG_MHC; 3.
 DR Hypothetical protein.
 KM Hypothetical protein.
 SQ SEQUENCE 618 AA; 67758 MW; 96BBD4C7C696E0A6 CRC64;

Query Match 16.1%; Score 389.5; DB 4; Length 618;
 Best Local Similarity 23.9%; Pred. No. 6.2e-24;
 Matches 116; Conservative 77; Mismatches 167; Indels 125; Gaps 16;

```

QY 36 TVELCTTASQ--KKSIOFHMKNSNOIKILGNQGSFLTGPSKLNDRADSRSSIMDQGNFP 93
DB 167 SVAAGCCLADFLPDSITFSWKYKNNSDISSTRG-----FP 201
QY 94 LIINKLKIEDSDTYICEVEDQKEVQLLVGLTANSPTLL-----QGQSILTT 142
DB 202 SVLRGCKXAATSOVLPSKD-----VMQGTDEHVCKVQHNGKKNVPLPI 250
QY 143 LESPSSPSVQCR-----SPR-----GKNIQGGKTLVSQ 174
DB 251 AELPKVSVFVPPRDGFNGNPKRSKLICQATGSPROIQVSWLRGKQVSGVTTDQVQA 310
QY 175 ELQDSG-----TWCTVLAQNKVKEFKIDIVPCAPAPKSCDKT 213
DB 311 EAKESGPTTYKKTSTLTIKESDMLSGMFTCRV--DHRGLTFQONASMCVDDQTAIR- 367
QY 214 HTCELLGGSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFMVVDGVEVHNAAK 273
DB 368 -----VFALPPS-PASIFLTSTKTLCLVLTDLTTYD-SVLTISWRONGEAVKTH 414
QY 274 TKPREQYNSTYRVSVLTVLHODWLNKKEYCKKSNKALPAPIKTISSAKQGP-REPO 332
DB 415 TNISSEHNATFSAVGSASICEDDMNSGSRFTCTVTHDLPSPFKQIISRKGVALHRPD 474
QY 333 VYTLPPSRDEL-T-KNOVSLTCLVKGFPYPSDIAVWESNGQP--ENNYKTTPEVLD--SDG 387
DB 475 VYLLPPAREQLNLRSAITITCLVTGFSPADVFVQMORGQPLSPKQVTSAPMEPEQAPG 534
QY 388 SFELYSKLTVDKSRMOQGNVSCSVMEHALHNHYTQKSLSLSPGLQDCECAEQDDEL 447
DB 535 RYFAHSILTVSEENGTGETYTCVAHEALPNRYETRTVDS-----TEGEVSADDEGFE 589
QY 448 GLWTT 452
DB 590 NLMAT 594

```

RESULT 42

O77601 PRELIMINARY; PRT; 86 AA.
 AC O77601;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3) (Fragment).
 OS Lophocobus albigena albigena.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopitheidae; Lophocobus.
 OX NCBI_TaxID=75568;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98320644; Pubmed=9656488;
 RA Harris E.B.; Disotelli T.R.;
 RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (primates: Papionini).";
 RL Mol. Biol. Evol. 15:892-900(1998).
 CC -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).
 CC -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).

```

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057391; AAC25135.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR InterPro: IPR007110; I9_1like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT NON TER <1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86 86
SQ SEQUENCE 86 AA; 9463 MW; 2BD97A8464FE9AB CRC64;

Query Match 16.1%; Score 388; DB 6; Length 86;
Best Local Similarity 87.1%; Pred. No. 5.6e-25;
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 107 YICEVDQKEVQLVFGLTANSPTHLTGQSLTLTLESPGSSPSVQCRSPRGNIOGG 166
Db 1 YICEVDKKEVELLVFGLTANSPTHLTGQSLTLTLESPPTSPSVKCRSPRGNIOVG 60

RESULT 43
Q9BU10 PRELIMINARY; PRT; 597 AA.
ID Q9BU10;
AC Q9BU10;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
DE Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strusberg R.;
RU Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
DR HSSP: P01825; 7FAB.
DR InterPro: IPR007110; I9_1like.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; I9_5.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE; 5.
DR PROSITE: PS50290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 597 AA; 65274 MW; 2DAFA8FB7E055851 CRC64;

Query Match 15.9%; Score 385; DB 4; Length 597;
Best Local Similarity 24.9%; Pred. No. 1.4e-23;
Matches 114; Conservative 73; Mismatches 164; Indels 106; Gaps 15;

Qy 36 TVELCTASQ--KKSIOFHMKNSQIKILNGSFLTGKPSKLNDRADSRSLMDQGNFP 93
Db 167 SVAAGCLADFLPDSTTSWTKYKNSDISSTRG-----FP 201

Qy 94 LIINKLKIEDSDTYICEVD---QKEVQLVFGLTANSPTHLTGQSLTLTLESPGS 149
Db 202 SVLRGKVAATQVLLPSKDVMQGTDEHYVCKVQHPNGNKEKNV---PLPVIALPFRV 257

Qy 150 SPVSQCR-----SPR-----GKNIQGGKTLVSQLELDQSG- 180
Db 258 SVFVPPRQDGFNGNPKRSKLICQATGFSFROIQVSWLRBGKQVGSVTTDQVQAEAKESGP 317

```

```

Qy 181 -----TWCTVLONOKVEKIDIVPCAPPEPKSCDKTHTCPPEL 220
Db 318 TTYKATSTLTIKESDWLSQSFNCRV--DHKGLTQQAASSVCVDDQDTAIR----- 367

Qy 221 GGPVSFLPPPKRQDTLMSRTPEVTCVVVDVSHEDPEVKFMNYVDGVEVHNAKTRPREQ 280
Db 368 ----VFALPPG-FASIFLTQSKTLCLVTLDTTYD-SVTISWTRNGEAVKTHNISSH 421

Qy 281 YNSTRRVSVLVTHQMDLNGKEYKKCKSNKALPAPIKTIISKAGQ-REPOVYTLTPS 339
Db 422 PRATSAVGASISICEDMNSGERFCTVTHDLPSPKQTTISRPKGVALLHRPDVYLLPPA 481

Qy 340 RDELTF-KQVSLTCLVKGFPSPDIQVWESNGQF--ENNYKTPPLVD--SDGSFELYSK 394
Db 482 REQLNLRSAITITCLVTGFSFADVGVQMQRGQPLSPKRYTSAFMPERQARFYAFASI 541

Qy 395 LTVDSRRMQQGNVPSGVNHEALHNHYTQKSLSLSPG 431
Db 542 LTVSEEWNTGRTYTCVVAHEALPRTVTRTVDKSTG 578

RESULT 44
Q9BOB8 PRELIMINARY; PRT; 597 AA.
ID Q9BOB8;
AC Q9BOB8;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
DE Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle; and Lymph;
RA Strusberg R.;
RU Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC006180; AAH06180.1; -.
DR EMBL: BC001872; AAH01872.1; -.
DR HSSP: P01825; 7FAB.
DR InterPro: IPR007110; I9_1like.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; I9_5.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE; 5.
DR PROSITE: PS50290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 597 AA; 65300 MW; 2DAFAD50A6375851 CRC64;

Query Match 15.9%; Score 385; DB 4; Length 597;
Best Local Similarity 24.9%; Pred. No. 1.4e-23;
Matches 114; Conservative 73; Mismatches 164; Indels 106; Gaps 15;

Qy 36 TVELCTASQ--KKSIOFHMKNSQIKILNGSFLTGKPSKLNDRADSRSLMDQGNFP 93
Db 167 SVAAGCLADFLPDSTTSWTKYKNSDISSTRG-----FP 201

Qy 94 LIINKLKIEDSDTYICEVD---QKEVQLVFGLTANSPTHLTGQSLTLTLESPGS 149
Db 202 SVLRGKVAATQVLLPSKDVMQGTDEHYVCKVQHPNGNKEKNV---PLPVIALPFRV 257

Qy 150 SPVSQCR-----SPR-----GKNIQGGKTLVSQLELDQSG- 180
Db 258 SVFVPPRQDGFNGNPKRSKLICQATGFSFROIQVSWLRBGKQVGSVTTDQVQAEAKESGP 317

Qy 181 -----TWCTVLONOKVEKIDIVPCAPPEPKSCDKTHTCPPEL 220
Db 318 TTYKATSTLTIKESDWLSQSFNCRV--DHKGLTQQAASSVCVDDQDTAIR----- 367

Qy 221 GGPVSFLPPPKRQDTLMSRTPEVTCVVVDVSHEDPEVKFMNYVDGVEVHNAKTRPREQ 280

```

```

Db 368 ---VFAIPPS-FASIFLTSTKLTCLVTLDTLTYD-SVTISWTRONGEAVKTHNISESH 421
QY 281 YNSTRYVSVLTVLHQMUNGKEYCKYKNKLPAPIKETISKAGQP-REPOVYTLPPS 339
Db 422 PNATSAANGEASIGCDMDNGSERFTCYTHDLPSPKQTSIRPGVALHHPDYLLPRA 481
QY 340 RDELTL-KNOVSLTCLVKGFPYPSDIADVEMESNGQP--ENNYKTPPVLD-SDGSFFLYSK 394
Db 482 REQLNLRASATITCLVTGFSFADVPVQWMQGOPLSPKRYTSAPMEPQAPGRYFAHSI 541
QY 395 LTVDKSRWQGVNFCSCVMHEALHNHYTKSLSPG 431
Db 542 LTVSEEMNTGETYTCVAHREALPNRTERTVDKSTG 578

```

RESULT 45

```

096BB9 PRELIMINARY; PRT; 597 AA.
AC 096BB9;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell;
RA Strauberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC015760; AAH15760.1; -
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS00835; IG LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 597 AA; 65039 MW; 4FCA3AD8E6CE263D9 CRC64;

```

Query Match 15.9%; Score 385; DB 4; Length 597;
 Best Local Similarity 24.9%; Pred. No. 1.4e-23;
 Matches 114; Conservative 73; Mismatches 164; Indels 106; Gaps 15;

```

QY 36 TVELTCTTSQ--KSIQTHMKNSNOIKILGNQGSFLTGPKSLNDRADSRSLMDQGNP 93
Db 167 SVAVCCLAQDLPLDSITPSMKYKNNSDISSTRG-----FP 201
QY 94 LIKKLKIEDSDPTVYCEVED---QKEEYQLLVFGLTANSDTHLLQGSLTLTLESPPGS 149
Db 202 SVLRGGKTAATSQVLLPSKDVWGQTDENVVCKVQHPNKKENV---PLVIAELPKV 257
QY 150 SPVSQCR-----SPR-----GKNIOGKTLVSQLELDQSG- 180
Db 258 SVFVPRDQFGNPKRSKILICATGFSRQIQVSLRKGQVGSVTTDQVQAEKESGP 317
QY 181 -----TWTCVYLONQKKVEXKIDIVCPAPPEPSCKTHTCPELL 220
Db 318 TTYKVTSTLTIKESDMLSQSMFTCRV--DHRLGTFQNASMVCVDDDTAIR----- 367
QY 221 GGPVFLPPPKDITLMSRPEVTCVAVDVSHEDPEKFMVYDVGVVNAKTRPREQ 280
Db 368 ---VFAIPPS-FASIFLTSTKLTCLVTLDTLTYD-SVTISWTRONGEAVKTHNISESH 421
QY 281 YNSTRYVSVLTVLHQMUNGKEYCKYKNKLPAPIKETISKAGQP-REPOVYTLPPS 339
Db 422 PNATSAANGEASIGCDMDNGSERFTCYTHDLPSPKQTSIRPGVALHHPDYLLPRA 481
QY 340 RDELTL-KNOVSLTCLVKGFPYPSDIADVEMESNGQP--ENNYKTPPVLD-SDGSFFLYSK 394

```

```

Db 482 REQLNLRASATITCLVTGFSFADVPVQWMQGOPLSPKRYTSAPMEPQAPGRYFAHSI 541
QY 395 LTVDKSRWQGVNFCSCVMHEALHNHYTKSLSPG 431
Db 542 LTVSEEMNTGETYTCVAHREALPNRTERTVDKSTG 578

```

RESULT 46

```

077600 PRELIMINARY; PRT; 86 AA.
AC 077600;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Lophoceros aterrimus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Lophoceros.
OX NCBI_Taxid=75566;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98320644; Pubmed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -I- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
SIMILARITY).
CC -I- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057390; AAC25134.1; -
DR HSSP; P01730; LCDY.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR InterPro; IPR007110; IG-1-like.
KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON TER 1 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86 86
SQ SEQUENCE 86 AA; 9476 MW; A9D97A88464FE93BE CRC64;

```

Query Match 15.9%; Score 383; DB 6; Length 86;
 Best Local Similarity 85.9%; Pred. No. 1.5e-24;
 Matches 73; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

```

QY 107 YICEVEQKEVQLVPLGLTANSDTHLLQGSLTLTLESPPGSPVSQCRSPGKNIQGG 166
Db 1 YICEVEDKKEVELLVFGLTANSDTHLLBEGSLTLTLESPPGSPVSQCRSPGKNIQVG 60
QY 167 KTLVSQLELDQSGITCTVLYLONK 191
Db 61 RTLSVPLERQDSGTWCNVSDQK 85

```

RESULT 47

```

077MT6 PRELIMINARY; PRT; 614 AA.
ID 077MT6;
AC 077MT6;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Rachekeno L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stepleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millihy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Rahy J., Helton E., Kettelman W., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
Krzyniaki M.J., Skolala U., Smalls D.E., Scherch A., Schein J.E.,
Jones S.J., Marra M.A.,
"Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RA Strausberg R.,
Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053409; AAH53409.1; -
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 67746 MW; 839BAF3B8D24F89 CRC64;

Query Match 15.8%; Score 381; DB 11; Length 614;
Best Local Similarity 23.3%; Pred. No. 3.2e-23;
Matches 139; Conservative 77; Mismatches 171; Indels 210; Gaps 25;

QY 30 LKKKGGTVELTCTAGSKSIQFHMK-----SNQKILGN-----OGSFLITGCP 73
DB 30 LVKPGASVIVISICAS-GVAFSSMMWVQKRPCKLEWIGRVYPGDGDITNYNGK--KKG 86
QY 74 SKLN-DRADSRSLMDQGNFPLIKLKIEDSDTYCEVEDQKEVQLVFLGTANSPTN 132
DB 87 ATLTLADKSSSTAYMQ-----LSLTSEDSAVYVC-ARDYGSYFRAYMG----- 129
QY 133 LIQGSGLVLTLESP--PGSSPSVQCRSP-RGKNI----- 163
DB 130 --QGLTVVASQSFPNVPFLVSCSPSLDKNLVAMGLARDFLPSTISFTWYQNNAE 187
QY 164 -----QGGKTLVSQLELQDSGT-----WTCTVLQONQKVEFKIDIVPCA 204
DB 188 VIQGITFPTLTGTGKYLATISQVLLSPKSLSEGSDEYLVCKIHGGKNDLH---VPIPA 244
QY 205 -----PEPKS-----CDKTHTCPELL----- 220
DB 245 VAEANPNVNVFVPRDGSFGSPAPRKSLICEATINFPKPIYVSMKDGKLVSESGFTTDPV 304
QY 221 -----GAPS-----V 225
DB 305 TIENKSGTPQYKVISLTITSEIDMLNVYTCRVDRHGLFLKLVNSVSCASPSTDIAT 364
QY 226 FLPPPKKDTLMSRTPEVTCVVDVSHEDPEYKFMWYDGVNNAKTPREEDQNSY 285
DB 365 FTTPSPFAD-IFLSKSNLITCLVSNLATYE-PLNLSMASQSGEPLETKIKISHPNNGF 422
QY 286 RVVSVTLVHODMLNGKEYKCKVSNKALPAPIEKTISKAGQPR-----POVYTLPPSR 340
DB 423 SAKGVASVCEVEDMNRKKEFVCTVTHRDLPSPQKKFISK-----PNEVHKRPRAVYLLPPAR 478
QY 341 DELT-KNOVSLTCLVKGFFPSDIAVWESNGQ--PENNYKTPPVLD--SDGSFELYSK 395
DB 479 EQLNLRESATVITCLVKGFSFPADISVQWLQRGQLLPQEKVVTSAFMPDEPAPGFYFTHTSL 538

QY 396 TVDKSRMOQGVNFSQSVNHEALNHNHYTKOISLSIPGLQDLETCAEAOGEDELGWT 452
DB 539 TVTEBNNNGSEITTCVSHALPHLVTERTYDKS-----TEGEVNAEEBGEENLMTT 590

RESULT 48
ID O8WUX4 PRELIMINARY; PRT; 588 AA.
AC O8WUX4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DR 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
OS Hypothetical protein.
OC Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.,
Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019235; AAH19235.1; -
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_5.
DR SMART; SM00406; IG_1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 588 AA; 64438 MW; F6C0DABD2B39FD7 CRC64;

Query Match 15.7%; Score 380; DB 4; Length 588;
Best Local Similarity 24.8%; Pred. No. 3.6e-23;
Matches 113; Conservative 73; Mismatches 163; Indels 106; Gaps 15;

QY 36 TVELTCTASQ--KKSIOFHMKNSQIKLNGSGSLTGTGPKLNDRADSRSLMDQGNF 93
DB 167 SVAVGCGLADPLPSITGSMYKKNSSDISTRG-----FP 201
QY 94 LIKKLKIEDSDTYCEVED-----QKEVQLVFLGTANSPTNLIQGSGLVLTLESPPGS 149
DB 202 SVLRGKKAATISQVLLPDKDWMQGTDEHYVCKVQHPNKNKKNV---PLVYIAELPPKV 257
QY 150 SPSVQCR-----SPR-----GKNIQGGKTLVSQLELQDSG- 180
DB 258 SVFVPRDGFQGNPKRKSLICQATGFSPRQIOWSLRGGQVSGVTTDQVQAEKESGP 317
QY 181 -----TWCTVLQONQKVEFKIDIVCPAPPEKSCDKTHTCPELL 220
DB 318 TTYKVTSTLTIKESDMLSQSMFTCRV--DHRGLTFQGNASSMCVDDPTAIR----- 367
QY 221 GAPSFLPPPKKDTLMSRTPEVTCVVDVSHEDPEYKFMWYDGVNNAKTPREEDQ 280
DB 368 ----VFAIPPS-FASIFLTGSKTLCLVTDLTITD-SYTIISWTRONGEAVVTHNLSIS 421
QY 281 YNSTYRVSVTLVHODMLNGKEYKCKVSNKALPAPIEKTISKAGQPR-REPOVYTLPPS 339
DB 422 PNATPSAVGASISICDDNNSGERFCTVTHRDLPSPKQTLISRPQVNLHNRDYLPPA 481
QY 340 RDELT-KNOVSLTCLVKGFFPSDIAVWESNGQ--ENNYKTPPVLD--SDGSFELYSK 394
DB 482 REQLNLRESATVITCLVTFSPADVFWQMOGQPLSPKPYTSAFMPDEPAPGFYFAHSI 541
QY 395 LTVDKSRMOQGVNFSQSVNHEALNHNHYTKOISLS 429
DB 542 LTVSEENNTGETTTCVVAHALPKNRYTERIVDVS 576

RESULT 49
ID O8VCX7 PRELIMINARY; PRT; 613 AA.

AC Q8VCX7; 08Vcx7; 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN IGH-6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strusberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC018315; AAH18315.1; -
 DR MGD; MGI:96448; Igh-6; -
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 5.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 5.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Hypothetical protein.
 SQ SEQUENCE 613 AA; 67855 MW; 41A9384DD4C22862 CRC64;

Query Match 15.5%; Score 373.5; DB 11; Length 613;
 Best Local Similarity 25.2%; Pred. No. 1,4e-22;
 Matches 121; Conservative 87; Mismatches 157; Indels 115; Gaps 23;

QY 37 VELTCTAQ--KKSIOFHKNKSNQIKILGNQSGFLTKGPSKLNDRADSRRLMDQNF-- 92
 DB 161 VAMGLADFLPSTISFTWNYQNTREVIGIRTPF-----LRTGSKVLA 205
 QY 93 ---PLIINKLKIEDDT-ICEVEQKEVOLVGLTANSPTLH---LQGSITLLE 144
 DB 206 TSQVLLSPKSLIEGDEYLVCIKH-----YG-GKNRDLHVPIDAVEMNPVNVF 254
 QY 145 SPP-----GSSP-----SVOCR-----SPR-----GKNIOG----- 166
 DB 225 VPRPGFGSPAPRKSKLICETATPFPKPIYVSKLQGLVSGFTTDTVTIENKSTPOT 314
 QY 167 ---KTLVSQLELDOSGTWCTVLONQKVEFKIDIVPCAPAEKSCDKHTCCELLG 222
 DB 315 YKIVISTLTSEIDMLNLYTCRV--DHRGLTFKLVGSTCAASPST-----DIL-- 362
 QY 223 PSVFLPFPKPDLTLMISRTPEVTCVVVDVSHEDPEVKENMYVDGVEVNNATKPREQYN 282
 DB 363 --TFIIPSPFAD-IFLSKSNALTCVSNLATYE-TLNI SMSQSGEPLETIKIMESHNP 418
 QY 283 STYRVVSVLTVDHOMLNGKEYKCKVSNKALPAPIEKTISKAKGQPRE-----POVYTL 337
 DB 419 GTFSAKGAVASVEVEEMNKRKEFCVTHRDLPSPQKTIK-----PNEVHKHPAVYLLP 474
 QY 338 PSRDEL-T-KNOVSLTCLVKGFPYSDIAVWESNGQ--PENNYKTPPVLD--SDGSF 392
 DB 475 PARQQLNRESATVTCVVGFSFADISVQMLQRLPOEKVITSAPEPBPAPGFYFT 534
 QY 333 SKLTVDKSRMOQGNVFSQSVMEALAHNYTQKSLSLSGQLDDEFCAADGELDLMTT 452
 DB 535 SILTVTEEBWMSGETTYTCVVGHEALPHLVTERTVKS-----TEGEVNAEBSGEMLM 589

RESULT 50

Q86TT1 PRELIMINARY; PRT; 375 AA.
 AC Q86TT1;
 DT 01-JUN-2003 (Tremblrel. 24, Created)
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Human full-length cDNA clone CS0DD006102 of neuroblastoma of Homo sapiens (Human).

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RA Genoscope;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RA Li W.B., Gruber C., Jaesee J., Polayes D.;
 RT "Full-length cDNA libraries and normalization."
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX161420; CAD61894.1; -
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00407; IGV; 3.
 DR PROSITE; PS50835; IG_LIKE; 3.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Plasmid.
 SQ SEQUENCE 375 AA; 41272 MW; 7ACD1AF4399C5EFE CRC64;

Query Match 15.2%; Score 366.5; DB 4; Length 375;
 Best Local Similarity 28.6%; Pred. No. 2.7e-22;
 Matches 98; Conservative 59; Mismatches 125; Indels 61; Gaps 12;

QY 139 LTTLESPPSSPSVOCR-----SPR-----GKNIOGKTL 170
 DB 25 LPVIAELPDKSVYVPRDGFNGNPKSKLICATGSPFQIQVSWLREGKQVSGVTTD 84
 QY 171 VSQLELDSC-----TWCTVLD---NOKVEFKIDIVPCAPAEKSCDKHTCCELL 220
 DB 85 QVQAEAKESGPTTYKVTSTLTIKESDWLSQSMFTCRVDHRLTFQGNASS-----M 135
 QY 221 GSP-----VFLPFPKPDLTLMISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVNNAT 274
 DB 136 CGPQDPAIKVFAIPSP-FASIFLTSTKTLCTLVTDLTLD-SVTISWTQNGEAVTHT 193
 QY 275 KPREQYNSTYRVVSVLTVDHOMLNGKEYKCKVSNKALPAPIEKTISKAKGQ-REP 333
 DB 194 NISSHNATFSAVGEASICEDDMNSGERFTCVTHDLPSPKQFISRPKGVALHRPDV 253
 QY 334 YTLPPSRDEL-T-KNOVSLTCLVKGFPYSDIAVWESNGQ--ENNYKTPPVLD--SD 388
 DB 254 YLPPAREQLNRESATITCLVGFSPADVFVQMGQPLSPKQVTSAPMPBPAPGR 313
 QY 389 FFLSKLTVDKSRMOQGNVFSQSVMEALAHNYTQKSLSPG 431
 DB 314 YFASILTVEEBWMSGETTYTCVVGHEALPHLVTERTVDSSTG 356

RESULT 51

Q13969 PRELIMINARY; PRT; 71 AA.
 AC Q13969;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE CD4 protein (Fragment).
 GN CD4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zverev V.V., Blinov V.M., Nedospasov S.A.;
 DT "Splice-mediated insertion of antisense and sense Alu repeats in human

RT CD4 gene: identification of three exons of CD4 mRNA."
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95407135; PubMed=7676667;
RA Zverev V.V., Sidorov A.V., Nedopasov S.A., Maliushova V.V.,
RA Udalova I.A., Andshapartide O.G., Blinov V.M.;
RT "[Nucleotide sequence of two exons of the human T-lymphocyte CD4
RT receptor gene]."
RL Vopr. Virusol. 40:100-102(1995).
DR EMBL; X87579; CAA60883.1; -
DR EMBL; S79267; AAB35273.1; -
DR PIR; I60082; I60082.
DR HSSP; P01730; 1CDY.
DR GO; GO:0004872; F;receptor activity; IEA.
DR Interpro; IPR007110; IG_1like.
DR PROSITE; PS50835; IG_LIKE; 1.
KM Receptor.
FT NON TER
SQ SEQUENCE 71 AA; 7644 MW; A5C9D84816135C86 CRC64;

Query Match 14.8%; Score 357; DB 4; Length 71;
Best Local Similarity 98.6%; Pred. No. 1.7e-22;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NMRGVFRLLLVQLALPPAATQGNKVVLGKKGDVVELTCTASOKKSIQFMKNSNOK 60
DB 1 NMRGVFRLLLVQLALPPAATQGNKVVLGKKGDVVELTCTASOKKSIQFMKNSNOK 60

QY 61 ILGNQSFLLTK 71
DB 61 ILGNQSFLLTK 71

RESULT 52
Q7TOR1 PRELIMINARY; PRT; 587 AA.
ID Q7TOR1
AC Q7TOR1
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Straubeberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT Initiative."
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uediri T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abrahams R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,
RA Villalón D.K., Muray D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalits D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Klein S., Straubeberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC056078; AAB6078.1; -
KM Hypothetical protein.
SQ SEQUENCE 587 AA; 64933 MW; 646B5FD0FA414E8D CRC64;

Query Match 14.3%; Score 345.5; DB 13; Length 587;
Best Local Similarity 25.8%; Pred. No. 2.8e-20;
Matches 119; Conservative 77; Mismatches 189; Indels 77; Gaps 18;

QY 31 GKKGDTVELTCTASQ-KKSIQFMKNSNOKILGNQSF---LTKPSPKLNDRADSRR 84
DB 149 GSSMDPVTIGCIADKFLPBITISFTWGDKNASYSYGLKSYKVMQSGTYSASQVNVAS 208

QY 85 SLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEVQ-----LLV 122
DB 209 AAWD--NIBQFYCNKHLDT--IKSVELKDPVPEVEKPVVSIHPSPKDALALNESLFI 263

QY 123 FGLTAN-SDTHLLQGSSTLTLESPPGSSPVQCRSP---RGKNIQGGKTLVSQLEIAD 178
DB 264 VCLATNFTPTHT-----VIKWLKNGNQTEGVRVPEPDKRGVLAISYLSITKEDWL 318

QY 179 SGTWTCTVLQNKQKVEKIDIVPCAPPEKSCDKTHTCPRELIGPS--VLEPPPKDQL 236
DB 319 DTVLSCV-----EHAESGSLQEKMSKSLKMDTPTPTSIQVITRPS-LESI 366

QY 237 MISRTPEVTCVVVDVSHEDPEVKFMY--VDGVEVNAKTKREEDYNS--TYRVSVLTV 293
DB 367 FEKSKATVTLCLVSNANSEDLRSISWFKSGQEIPLTELGDALINNRITSVAGCTTV 426

QY 294 LHQDLINKEVYCKVSNALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTLC 352
DB 427 CADENNNNK-FVCKVEHHLASMKVPLFKKEGVENTSVYFPPLELSRETRATLTL 485

QY 353 LVKGYPSYDIAYEW--ESNGQPENNYKTP-----PYLSDGSGFFLYSKLTVDKSRMQQ 404
DB 486 LVKGFSPSEIIFVKWILHKEAIVPKQYINTSINDELLPRGQKSGKFFLYSLHTIDIKWDA 545

QY 405 GNVPSCSYMHEALNHYTQKSLSPG-----LQDENC 438
DB 546 GDSFSCVGHSLPLQLTORSIDKSSGKPTNVNVLVLDTC 587

RESULT 53
Q7Z379 PRELIMINARY; PRT; 478 AA.
ID Q7Z379
AC Q7Z379
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZP686K04218 (Fragment).
GN DKFZP686K04218.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human rectum tumor;
RA Bloeker H., Boecher M., Mewes H.W., Weil B., Amid C., Oanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX538066; CAD97996.1; -


```
QY 62 LKNGSF-LTKGPSKLNDRADSRSL--WDQGNFLLIKNLIKEDSDTYICEVEDQKEE 117
DB 64 LKMGKVNIEGESYVADPFKGFAPSLSTASTIHLQJNMLNKNEBTATYPCARSDYD 123
QY 118 VOLLVEGLTANSDTHLLQGOSLTLTLES-----PGSS-----PS 152
DB 124 IYAM-----DWGQGTSTVVSSESARNPITYPLLPALSSDPVITGLIHDPFS 174
QY 133 VOCRSPRGKN-----IQGKTLVSQI-----ELDQSGTWCTYLQNOKKYE 194
DB 175 GTMNATWVGSGKDITTVNFPALASGGRYTMSSQLTPAVECPREGSVKCSVOHSPNQ 234
QY 195 FKIDIVCPAPPEKSCDKHTHCPBELLGGSVFLPPPKDPTMISTREPVTCVVVDVSH 254
DB 235 -ELD-VNCGPTPPPTTPSCQ-----PSLSLQRPALBD-LILGSDASITCTLLGNLRP 286
QY 255 DPEVKFNMYVDGVEVNAKTKPREBOYNST--YRVSVLTVLHOPMLNGEKYCKVSNK 311
DB 287 EGAV-FTW-----EPSTGDAVQKAVQNSCGCYSSVLPQCAERMSNGASFCTVTHP 340
QY 312 ALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVWESNG 370
DB 341 E-SGTLTGTAIVTVTPPPQVHLPPPESEELALNELSLTCLVFAFNKEVYLVWMLGN 399
QY 371 Q---PENNYKTTPPVLDSG--SFFLYSKLTVDKSRMOQGNFSCSVMEALHNHYTOKS 425
DB 400 EELSPESYVFEPLKEPGEATTVLTVSVLRVSAETWKQGDQYSCMVGHEALPMNFTQKT 459
QY 426 LSLSPG 431
DB 460 IDRLSG 465
```

```
RESULT 56
Q99KA4 PRELIMINARY; PRT; 487 AA.
ID 099KA4
AC 099KA4
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC004786; AAH04786.1; -.
DR HSSP; F01810; 2FBU.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 487 AA; 52554 MW; 7DC8E96DB333077B CRC64;
```

```
Query Match 13.0%; Score 314.5; DB 11; Length 487;
Best Local Similarity 26.2%; Pred. No. 8.6e-18;
Matches 132; Conservative 76; Mismatches 188; Indels 107; Gaps 25;

QY 1 NNRGVFRRLLVL--QLALPAAATQGNKVVLGKKDGVETLCTAS--QKSGIOFHW-- 53
DB 1 NMFGLSLFLVLVLKVGQCEVQLVESGGGLV--KKGSLKLSGCAASGFTSSSYAMSVKR 57
QY 54 ----KNSNOIKILGNGSFLLTKGPSKLNDRADSRSLDQGNFPLIKLIKIEDSDTYIC 109
DB 56 QTPKRLKLEWVATISDGSY-TYYPDVVKGRFTISRDN-AKNMLYIQMHLKSKEDTAMVYC 115
```

```
QY 110 EVEDQKEVQLVFLGTANSDTHLLQGOSLTLTLESPPGSSPSV-QCRSPR----- 159
DB 116 ----ARDMGSPYGGYRFD---YWGQTTITVSSSESARNPITYPLLPALSSDPVIT 167
QY 160 -----GKNIQ-----GKTLVSQI-----ELDQSGTW 183
DB 166 GCLIHDPFSGMTVNTWVGSGKDITTVNFPALASGGGYTMS-SQLTPAVECPREGSVK 226
QY 184 CTVLQNOKKVFEKIDI-----VPCPAPPEKSCDKHTHCPBELLGGSVFLPPPKDPTLM 237
DB 227 CSVQHSNAVQ-ELDVKCSGPPPCP-PCPSCH-----PSLSLQRPALBD-L 272
QY 238 ISRPEVPCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREBOYNST--YRVSVLTVL 294
DB 273 LGSDSLCTCTLLGNLRNPGAV-FTW-----EPSTGDAVQKAVQNSCGCYSSVSLPQC 326
QY 295 HQDWLNGEKYCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNQ-VSLTCL 353
DB 327 AERNMSGASFKCTVTHPESDT-LTGTLAKITVNTFPQVHLPPPESEELALNELVSLTCL 385
QY 354 VKGFYPSDIAVWESNGQ--PENNYKTTPPVLDSG--SFFLYSKLTVDKSRMOQGNV 408
DB 386 VRAFNPKEVLYVWMLGNELSPESYLVPEPLKEPGEATTVLTVSVLRVSAETWKQGDQY 445
QY 409 SCVMEALHNHYTOKSLSLSPG 431
DB 446 SCMVGHEALPMNFTQKTIDRLSG 468
```

```
RESULT 57
Q8VCX4 PRELIMINARY; PRT; 489 AA.
ID Q8VCX4
AC Q8VCX4
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS IGH-VJ558 OR A1893585.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue=Colon;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC018322; AAH18322.1; -.
DR MGI; MGI:96486; IGH-VJ558.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 489 AA; 53208 MW; CC85B1194DAFE2C CRC64;
```

```
Query Match 13.0%; Score 313.5; DB 11; Length 489;
Best Local Similarity 24.9%; Pred. No. 1.1e-17;
Matches 118; Conservative 70; Mismatches 181; Indels 105; Gaps 24;

QY 30 LKKKDYELTCTASQKKSIOF--HW--KNSNOIKILG--NQGSLTKGPSKLNDRADS 82
DB 30 LVKPGASVKLCKSGKSGYFSDYFIHWIKRSGQGLEWIGMFPNGS----GSIKFEKPKD 85
QY 83 RSLMDQGNFPLI--IKNLIKIEDSDTYICEVEDQKEVQLVFLGTANSDTHL-----LQ 135
DB 86 KATLTADSSSTTVYMDLSRLTSEDSAVYFCARHEDR-----GNYDGSILAMFYVW 134
QY 136 GQSLTLTLESPPGSSPSV-----QCRS-KNI-----QGGKTLV 171
```

```

Db 135 GGGTLVTSASBARREPTIYPLTFPQALSSDPVIGCLIHDFPSTGTMTVTKSGKIDITT 194
Qy 172 SOL--ELDSDGWTCTVQONQKVEFKIDYPCAPPEKSCDKTH-----TCBEL 219
Db 195 VNFPPALASGGGYTMS-----SQTLPVAVCEPGESVKSQVQDSDNPVQELNVNCGI 247
Qy 220 LGSPVFLFPPEKPK-----DTLMISRTPEVTCVVVDVSHEDPE-VKFNMYVDG 266
Db 248 CSPFTT---PPPPSCQPSLSIORPALBEDLLGSDASITCTLNGL--RPREGAVFTW---- 298
Qy 267 VEHNAKTKPREEOYNST--YRVVSVITVLHQMNLNGKEYKCKVSNKALPAPIEKTISK 323
Db 299 -EPSTGKAVQKQKAVQNSCGCYSSVLPGCALERNNSGASFCKYVTHHESDT-LTGITAK 356
Qy 324 AKGQREPOVYTLPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ--PENNYKTT 379
Db 357 VTVNTFPQVHLPPSSSELALNELVSLTCLVRANPKREVLRWMLHNGNEELSPESYLVE 416
Qy 380 PVLDSDCG--SFFLYSKLTVDRKRNQGNVFCSVMEALHNHYTKSLSPG 431
Db 417 FLKEGEGATTYLVTSVLRVSAETWKQGDQYSCMVGHALPMTFTOKTIDRLSG 470

```

RESULT 58

```

Q9NPP6 PRELIMINARY; PRT; 416 AA.
ID Q9NPP6 AC Q9NPP6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Immunoglobulin heavy chain variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN NM_011111.1
RP SEQUENCE FROM N.A.
RA Pluvinet R., Estivill X., Escarceller M., Sumoy L.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Aultray C., Ansoerge W., Ballabio A., Estivill X., Gibson K.,
RA Lehnach H., Pouscka A., Lundberg J.;
RT "The European IMAGE consortium for integrated Molecular analysis of
human gene transcripts.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL88978; CAB97534.1; -.
DR HSSP; P01789; IMCP.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
FT NON TER
SQ SEQUENCE 416 AA; 44786 MW; 8C41708B86A4687 CRC64;

```

Query Match 12.9%; Score 312; DB 4; Length 416;
 Best Local Similarity 27.8%; Pred. No. 1.1e-17;
 Matches 116; Conservative 61; Mismatches 163; Indels 78; Gaps 21;

```

Qy 78 DRADRRRLMD-----QGNFPLIKNKIEDSDTYIC-----EVEDQKEEYQL----- 120
Db 19 DYADSVKRFVYSRDTAKNSLSLQMSLRVEDTAVYCARYYGMDVWGQGTYYVSSAS 78
Qy 121 ----LVFGITANS DTH-----LLOG-----QSITLTLESPPGSSPSVQGRS-PRGKN 163
Db 79 PTSPKVFPLSDSTQDGNVAVACLVQGFPPQEPFLSVTWSF---SGQVTVARNPPPSDA 135
Qy 164 QGGKTLVSQLEL-----ODSGTWTCTVQONQKVEFKIDI-VPCPAPEKSCDKTHTCP 217

```

```

Db 136 SGLDYTTSQTLTPATQCPDGKSVTCHVKHYTNSQ---DVTWPCVPVPPPPC-----C- 186
Qy 218 ELGSPVFLFPPEKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKPR 277
Db 187 ---HPRLSLHRPALBD-LILGSBANLTCTGL-RDASGATFTWPSGK--SAVGCP 238
Qy 278 EEOYNSTYRVVSVITVLHQMNLNGKEYKCKVSNKALPAPIEKTISKAKGPREPOVYTL 337
Db 239 ERDLCGCVSVSVLPGCAQPMNGETFTCTAAHPELTKPILANTKS-GMTFREVHLFP 297
Qy 338 PSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ--PENNYKTTPVLD-SDG--SFFL 391
Db 298 PPSSELALNELVTLTCLARFSPADVLVRWLGOSQELPREKYLTVASROBPSQGTTFVAV 357
Qy 392 YSKLTVDRKRNQGNVFCSVMEALHNHYTKSLSPG-----LQDENC 438
Db 358 TSLIRVAEDWKKGDFTSCMVGHALPLAFTOKTIDRLAKPHTNVSVMAEVDGTC 415

```

RESULT 59

```

Q91WT1 PRELIMINARY; PRT; 481 AA.
ID Q91WT1 AC Q91WT1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN NM_011111.1
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013490; AAH13490.1; -.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 481 AA; 52105 MW; 97DF68D159463F65 CRC64;

```

Query Match 12.9%; Score 311; DB 11; Length 481;
 Best Local Similarity 24.7%; Pred. No. 1.7e-17;
 Matches 120; Conservative 67; Mismatches 187; Indels 112; Gaps 23;

```

Qy 15 QLALLPAATQGNKVVLGKGDVTELTCTAS--QKSIQFIMKNSNOIKIIGNO----- 66
Db 20 QVOLLQSGPE---LVKPAVSXISCKASGYFTSTYYHW---VKQPPGQLVWIGW 69
Qy 67 SFLTKGPSKLNDRAISRSLM---DQGNFPLIKNKIEDSDTYICEVEDQKEEYQLLVF 123
Db 70 IYPDGNTXKNKPKKGTITLADKSSSTAYMFLSLTSEDSAVYFC----- 115
Qy 124 GLTANSDTHLLQGSITLTLESPPGSSPSV-----QGRS-----PRG-KN 162
Db 116 -TRGGWAPFYWGQGTTLTVSSPAREPTIYPLTFPQALSSDPVIGCLIHDFPSTGTMN 174
Qy 163 I---QGGKTLVSQLEL-----ELDSDGWTCTVQONQKVEFKIDYPCAPPEKSCDKTH--- 214
Db 175 VTWKGSKDITTVVNFPPALASGGGYTMS-----SQTLPVAVCEPGESVKSQVQDSDN 227
Qy 215 -----TCBELGSPVFLFPPEKPK-----DTLMISRTPEVTCVVVDVSHED 255
Db 228 PVQELNVNCGICSPFTT---PPPPSCQPSLSIORPALBEDLLGSDASITCTLNGL--RD 282
Qy 256 PE-VKFNMYVDGVEVHNAKTKPREEOYNST--YRVVSVITVLHQMNLNGKEYKCKVSNK 311

```

```

Db      283 PEGAVFTW-----EPSTGKDAVQKAVQNSCGCYSSSVLPGCAERWNSGASFCKTVTHP 337
Qy      312 ALPAPIEKTISKAKQPRPPQVYTLPPSDELTKNQ-VSLTCLVKGFPYSDIAVEMESNG 370
Db      338 ESDT-LTGITAKVTVNTPPQVHLPPSEELALNELVSLTCLVAFNKEVLAHNEELVWMLGN 396
Qy      371 Q---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRWQGNVFGCSVMHEALHNHYTQKS 425
Db      397 EELSSESYLVFEPLEKPEGGATTYLVTSYLRVSAEIMKQGDQYSCMVGHEALPMNFTQKT 456
Qy      426 LSLSPG 431
Db      457 IDRLSG 462

RESULT 60
Q8K172 PRELIMINARY; PRT; 482 AA.
AC Q8K172;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to expressed sequence A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (Apr-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL, BC028249; AAH28249.1; -.
DR PIR; F33932; F33932.
DR PIR; PH1105; PH1105.
DR PIR; PH1108; PH1108.
DR PIR; PH1114; PH1114.
DR PIR; PH1118; PH1118.
DR PIR; PH1119; PH1119.
DR PIR; PH1125; PH1125.
DR PIR; PH1126; PH1126.
DR PIR; PH1128; PH1128.
DR PIR; PH1131; PH1131.
DR PIR; PH1134; PH1134.
DR PIR; PH1139; PH1139.
DR PIR; PH1142; PH1142.
DR PIR; PH1149; PH1149.
DR PIR; PH1150; PH1150.
DR PIR; PH1151; PH1151.
DR PIR; PH1152; PH1152.
DR PIR; PH1153; PH1153.
DR InterPro; IPR007110; IPR007110.
DR InterPro; IPR003597; IPR003597.
DR InterPro; IPR003006; IPR003006.
DR InterPro; IPR003596; IPR003596.
DR Pfam; PF00047; PF00047.
DR SMART; SM00407; SM00407.
DR SMART; SM00406; SM00406.
DR PROSITE; PS50835; PS50835.
DR PROSITE; PS00290; PS00290.
SQ SEQUENCE 482 AA; 52121 MW; A06F083E771D084 CRC64;

Query Match 12.9%; Score 311; DB 11; Length 482;
Best Local Similarity 25.0%; Pred. No. 1.7e-17;
Matches 117; Conservative 68; Mismatches 183; Indels 100; Gaps 22;
Qy 30 LGKQDPTVLTCTAS--QKSIQPFH---KNSNOIKIINGQSGFLTKGPSKLNDRADARR 84
Db 30 LVKPAQSVLSCAKAGYFTTSTYMHMVKQPRGRLGWIQRIDP--NSGQTKNEKFKSKA 87
Qy 85 SLW---DQGNPFLIINKLIEDSDTYICVEDQKEVQLVFGLTANSPDTHLLOGQSITL 141
Db 88 TLTVDKPSETAVMQSLTSEDSAVVYCYTRE-----GDYDAMDYWGQGNISV 133

```

```

Qy 142 TLSPSSPSV-----QCRS-----PRG-KNI---QGKTLVSQI--E 175
Db 134 TVSEPPAREPTIYLPPTPQALSDPVIIGCLIHDFPESGTVNWKSGKQITTVNPPPA 193
Qy 176 LQDSGTWTCTVLQNKQKVEFKIDIVPCAPAPKSCDKTH-----TCPELGGPSV 225
Db 194 LASGRYMS-----SQLTPAVECEGSGSVGSHDNPQVELVNVNPGICSPPTT 246
Qy 226 FLPPPKPK-----DTLMSRTPEVTCVVVDVSHEDPE-VKFNMYVDGVEVANA 272
Db 247 ---PPPPSCQPSLSLQRPALDELILIGSDASITCLNGL--RDPGAVFTW-----EPSTG 296
Qy 273 KTKPREBOYNST---YRVSVLYTLHQMVLNGKVEKCVSKALPAPIEKTISKAKQPR 329
Db 297 KDAVQKRAVQNSCGCYSSSVLPGCAERWNSGASFCKTVTHPESDT-LTGITAKVTVNTF 355
Qy 330 EPQVYTLPPSDELTKNQ-VSLTCLVKGFPYSDIAVEMESNGQ---PENNYKTPPVLD 385
Db 356 PPQVHLPPSEELALNELVSLTCLVAFNKEVLAHNEELSPESYLVFEPLEKPEG 415
Qy 386 DG--SFPLYSKLTVDKSRWQGNVFGCSVMHEALHNHYTQKSLSLSPG 431
Db 416 EGATTYLVTSYLRVSAEIMKQGDQYSCMVGHEALPMNFTQKTIDRLSG 463

RESULT 61
Q91X92 PRELIMINARY; PRT; 482 AA.
AC Q91X92;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (Jul-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL, BC011181; AAH11181.1; -.
DR PIR; F33932; F33932.
DR InterPro; IPR007110; IPR007110.
DR InterPro; IPR003006; IPR003006.
DR InterPro; IPR003596; IPR003596.
DR Pfam; PF00047; PF00047.
DR SMART; SM00406; SM00406.
DR PROSITE; PS50835; PS50835.
DR PROSITE; PS00290; PS00290.
KM Hypothetical protein.
SQ SEQUENCE 482 AA; 51865 MW; 312E01F9C1BC7F3C CRC64;

Query Match 12.7%; Score 307; DB 11; Length 482;
Best Local Similarity 25.0%; Pred. No. 3.6e-17;
Matches 123; Conservative 69; Mismatches 198; Indels 102; Gaps 24;
Qy 10 LLLVQLALPLPAATQGNKVVLGKKG-----DTVLTCTAS--QKSIQPFHMKNSNQIK 60
Db 4 LALLCLVTPSPSCL-SGVQLKESGPDVLVAPSQSLSTICTVSGFALSYAISWVRQPPGK 62
Qy 61 ILNGQSFUTKGPSKLNDRADARRSL-WDQGNPFLIINKLIEDSDT--YICEVEDQKEE 117
Db 63 GLEMLGVIMTQGYTNVNSALKSRSLISKDNKSGVFLKMSLQNTDFAHYC-ARDSNVE 121
Qy 118 VQLVFGLTANSPDTHLLOGQSITLTLSPSSPSV-----QCRS----- 157
Db 122 GAMDYV-----GGQTSVTVSSEPPAREPTIYLPPTPQALSDPVIIGCLIHDF 169
Qy 158 PRG-KNI---QGKTLVSQI--ELQDSGTWTCTVLQNKQKVEFKIDIVPCAPAPKSCD 211

```

Db 170 PSQTNVNTWKGSKDITTVNFPALASGGRYTM-----SQLTLPAVECPGEGSVKCS 222
Qy 212 KTH-----TCPELLGSPVFLFPKPK-----DTLMISRTPEVTCVV 249
Db 223 VQHDNSNPVQELNVNCPGICSPPT--PPPSGCPSLQRPALDELGLSDASTTCLN 279
Qy 250 DVSHEDPE-VKFNMYVDGEVHNNAKTKPREEQNST--YRVSVTLVHODMLNGKEYK 305
Db 280 GL--RDPGAVFTW-----EPSTGDAVQKAVQNSCCCYSSVSLPGCAERMSGASF 332
Qy 306 CKVSNKALPAPIEKTISKAGGPREPOVYTLPPSRDELTKNO-VSLTCLVKGFPYSDIAY 364
Db 333 CTVTPEESDT-LGTIAKYVTWTFPPQVHLPPPSSEELALNELVSLTCLVPAFNPKEVLV 391
Qy 365 EMESNGO--PENNYKTPPVLDSDG--SFLYSKLTVDKSRMOQGNVSCSVHMEALHN 419
Db 392 RMLHGNELSPESYVFEPLKEPGEGATTYLVTSVLAETWKGQDQYSCMVGEALPM 451
Qy 420 HYTKSLSLSPG 431
Db 452 NPTQKTIDRLSG 463

RESULT 62

Q29027 ID Q29027 PRELIMINARY; PRT; 99 AA.

DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CD4, allele 1 (Fragment).
OS Sue scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxId=9823;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN#1183; TISSUE=Blood;
RX MEDLINE=93329116; PubMed=8335933;
RA Gustafsson K., Germana S., Sundt T.M., Sachse D.H., Lequern C.;
RT "Extensive allelic polymorphism in an exposed region of the
RT minature. . . ."
RT J. Immunol. 151:1365-1370(1993).
DR EMBL; X65629; CAA46583.1; -.
DR PIR; I47131; S21461.
DR HSSP; P01730; 1CDY.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_1.
DR SMART; SMO0406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT 99
SQ SEQUENCE 99 AA; 11170 MW; 40BF080699C9F5D0C CRC64;

Query Match 12.7%; Score 305.5; DB 6; Length 99;
Best Local Similarity 60.2%; Pred. No. 5, 5e-18;
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 32 KKGDTVELTCTASQKSIQFHWKNSNQIKILGNQGSFL-TKGPSKLNDRASRLMDQ 90
Db 1 KAGDLAEIPCHSSQKKNLPFNWKNNSQTKILGHSFWHTASVTELTRLDSKKMMDHG 60
Qy 91 NPPLIKLKLEDTPTICEVEDKEEYQVLVFLGTAN 128
Db 61 SPPLIKLEVDTSIGYICEVEDKRIEVLQVLFRLTAS 98

RESULT 63

Q91207 ID Q91207 PRELIMINARY; PRT; 486 AA.
AC Q91207;
DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010324; AAH10324.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003066; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig_4.
DR SMART; SMO0406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 486 AA; 52682 MW; 4FEF835125DA870B CRC64;

Query Match 12.6%; Score 305; DB 11; Length 486;
Best Local Similarity 25.7%; Pred. No. 5, 4e-17;
Matches 129; Conservative 79; Mismatches 188; Indels 106; Gaps 25;

Qy 1 MNGVPRHLLVQLALPAATQGNKVIG----KKGDTVELTCTAS--QKSIQFHWK 54
Db 1 MNFGL---RLIFVLNLKQVCEVHLVSGGLVKFGSLKLSGVSGSFTSYDMKSV 56
Qy 55 NSNOIKILGNQGSFLTKGPSKLNDRASRL--WDQGNFPLIK--NLKIEDSPTYICEV 111
Db 57 RQPERLEWVAALITSGNYYPDNVKGRFTVSRDNAKYTLVYQMSLKSDEDTAMYC-- 114
Qy 112 EDQKEVQLLVF-GLTANSPTHLQGSGLTLTSPGSSPSV-QCRSPR----- 159
Db 115 --VREIPYIYSGSYEDS----WGQGTITVSSSARNPITYPLTLPALSSDPYIIG 167
Qy 160 -----GKNIQ-----GKTLISVSQL-----ELQDSGHWTC 184
Db 168 CLHDIYPSGTMNTWKGSKDITTVNFPALASGGGYTMS-SQLTLPAVECPGEGSVK 226
Qy 185 TVLQNKQKVEFKIDI-----VPCPAPEPKCDKTHCPBLGLGSPVFLFPKPKDTLMI 238
Db 227 SVQHDNSNAVQ-ELDVKSGSPPCP-PCPPSCH-----PSLSLQRPALSD-LLL 272
Qy 239 SRTPETVTCVVVDVSHEDPEVKFNMYVDGEVHNNAKTKPREEQNST--YRVSVTLVH 295
Db 273 GSDASLTCTLNGLRNPEGAV-FTW-----EPSTGKDAVQKAVQNSCCCYSSVSLPGCA 326
Qy 296 QDMVNGKEYCKVSNKALPAPIEKTISKAGGPREPOVYTLPPSRDELTKNO-VSLTCLV 354
Db 327 ERMNSGASFQCYVTHPESDT-LGTIAKYVTWTFPPQVHLPPPSSEELALNELVSLTCLV 385
Qy 355 KGFYPSDIAVWESNGO--PENNYKTPPVLDSDG--SFLYSKLTVDKSRMOQGNVFS 409
Db 386 RAIFPKKVLVRLHGNELSPESYVFEPLKEPGEGATTYLVTSVLAETWKGQDQYS 445
Qy 410 CSVMHEALHNHYQKSLSLSPG 431
Db 446 CMVGEHALPMNFTQKTIDRLSG 467

RESULT 64

Q8K0F2 ID Q8K0F2 PRELIMINARY; PRT; 488 AA.

AC Q8K0F2;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Similar to expressed sequence A1893385.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX NCBI_TaxId=10900;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RL Struhsberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC010798; AAH10798.1; -
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 2.
 KW Hypothetical protein.
 FT NON TER
 SQ SEQUENCE 480 AA; 51936 MW; 20B9234EEF2B41ED CRC64;

Query Match 12.4%; Score 300.5; DB 11; Length 480;
 Best Local Similarity 24.2%; Pred. No. 1.3e-16;
 Matches 121; Conservative 69; Mismatches 200; Indels 111; Gaps 23;

QY 2 NRGVFRHLVLLQLALPAATQGNKVVLGKKGVLTCTAS-----QKKS 48
 Db 1 NFGSLIFLVLLKGLCDVKLVESGGGLVKEGSLRLSCAAGPFGNSYMSWVRQTP 60
 QY 49 IOFHKNNSQIKILGNOSFLTKGPSKLNDRADSRSLMDQNFLLIKNLKIEDSDYI 108
 Db 61 KRLKH-----VATINSGLYATHYPSDSMKGRFTISRDN-AQNTVLLQMTSLNSEYVY 113
 QY 109 CEVEDQKEVOLVLEGLTANSDTHLLQGSLLTLLSPGSSPSV-----QCRS----- 157
 Db 114 CTRGD-----YWYFD-----VWGAGTIVTSSSEPARPTIYPLFPQALSSDPVI 158
 QY 158 -----PRG-KNI-----QGKTLVSQQL-ELDSGTWCTVQONQKVEFKIDYPC 202
 Db 159 IGLLHIDYFPGSTNMVWTKSGKIDITTVNPPALASGRGYTMS-----SGLTLPAVEC 211
 QY 203 PAPERKSCDKT-----TCPELLGSPSVLPFPPKPK-----DTLMISR 240
 Db 212 PEGSEVSKSVQHDSPVOELANNCPCIGSPRT-----PPSPSQPBLSIQRLBELLLIGS 268
 QY 241 TPEVTCVVVDVSHEDPE-VKFNMYVDGVEVHNNAKTRPEEOYNST---YRVSVLTVLHQ 296
 Db 269 DASICTLNGL--RDPEGAVFTW-----EPSTGDAVQKKAQNCSGCVSVSLPGCAE 321
 QY 297 DMLNKEKCKKSNKALPAPIEKTISKAKGQPREQVYTLPPSRDELTKNQ-VSLTCLVK 355
 Db 322 RMNSGASFRCTVTHESDT-LTGTLAKYVNTFPQVHLPPPESELALNELSLTCLVR 380
 QY 356 GFYSPDIAVEMESNGQ--PENNYKTPPVLDSDG--SFFLYSKLTVDKSRMOQGNVPS 410
 Db 381 AEPNPEVVLVRLHNGBELSPESYLVPFLPKEREGATYLVTSVLRSABEIKQDDQYSC 440
 QY 411 SVMHEALHNHYTQKSLSLSPG 431
 Db 441 MVEGHEALPMNFTQKTIDRLSLG 461
 RESULT 67
 ID Q96K68 PRELIMINARY; PRT; 494 AA.
 AC Q96K68;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein FL14473.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;
 RA Isegai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Maeuho Y.,
 RA Nimomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027379; BAB55072.1; -
 DR PIR; S21205; S21205.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 494 AA; 53088 MW; 9A1D7A8B5AEB4C0E CRC64;

Query Match 12.4%; Score 300.5; DB 4; Length 494;
 Best Local Similarity 25.3%; Pred. No. 1.3e-16;
 Matches 123; Conservative 71; Mismatches 193; Indels 99; Gaps 24;

QY 30 LGKKGVLTCTAS--QKKSIOFHW-----KNSNQIKILGNOSFL-----TKGPSKL 76
 Db 30 LVKFGSLRLSCAAGPFGNSYMSWVRQTP 89
 QY 77 NDRADSRSLMDQNFLLIKNLKIEDSDYICEVEDQKEVOLVLEGLTANSDTHLLQ 136
 Db 90 S-RDNKAGSLYLQWN-----SLRVDDTAVYYCARDCNGAL---CYGFSR-----WG 132
 QY 137 QSLTLLSPGSSPSV-----QCRSPKGN-----IQG--GKTLVSQQL----- 175
 Db 133 QGTLVYSSASPTSPKVPPLSLCSTPDGNNVIACTLVQGFPEPLSPVSESGQVYAR 192
 QY 176 -----LDPSG-----TWCTVQLONQKVEFKI-----DI-VPCPAPERKSCDK 212
 Db 193 NFPSPQASGDLITTSQTLTPATQCLAGSVCHVHYNPDSODVVPFVPSPTPTPS 252
 QY 213 THTCPELLAG--PSVLPFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDG 269
 Db 253 PSTPTPTSPSCPHRLSLHNPALPD-LILGSEANLTCTLGL-RDAGVFTWTPSSGK- 309
 QY 270 HNAKTREREGYNSTRVYVSVLVTLHODMLNGEYKCKVSNKALPAPIEKTISKAKQPR 329
 Db 310 -SAVQGPREDLCGCYSSVSLPGCAEPNMGKTFCTTAAYPESKTPLTATLSKS-GNTF 367
 QY 330 EPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNGQ--PENNYKTPPVLDS 385
 Db 368 RPEVHLPPPESELALNELVLTCLARGFSPKQVLYRMLOGSQBLPEKTLTMASROEPS 427
 QY 386 DG--SFFLYSKLTVDKSRMOQGNVPSCSVMHEALHNHYTQKSLSLSPG-----L 432
 Db 428 QGTTFAVTSILRVAADMKGDTFSCMGVGHALLPLAFTQKTIDRLAGKPTHVNVSVMA 487
 QY 433 QLDFTC 438
 Db 488 EVDQTC 493
 RESULT 68
 ID Q91WT3 PRELIMINARY; PRT; 481 AA.
 AC Q91WT3;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN Mus musculus (Mouse).
 OS

DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig_3.
 DR SMART; SM00409; Ig_3.
 DR SMART; SM00407; IGc1_3.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 2.
 DR Hypothetical protein.
 KW SEQUENCE 487 AA; 53019 MW; 31F2C893900A4D80 CRC64;

Query Match 12.3%; Score 297.5; DB 11; Length 487;
 Best Local Similarity 24.7%; Pred. No. 2,3e-16;
 Matches 118; Conservative 72; Mismatches 173; Indels 115; Gaps 24;

QY 30 LKKKDDVLELTCTASQKSIQFMKN-----SNQIKLGNQSFLLT-----K 71
 DB 30 LVQPGSGMKLSGAS-GFTFSNYMMWVQSPKGLFWVAEIRLSN--NYATHAESVK 86
 QY 72 GPSKLNDRADRSRLMDGNFPLIKNLKIEDSDTYICEVEDQKEVQLVPLGFLANSDT 131
 DB 87 GRFTTS-RDSSKSVYLQNN-----NLAEPTYGIYCTRR-----GYGDPWY 128
 QY 132 HLLQGQSLTLTLESPGSSPSV-----QCRS-----PRG-KNI---QCGX 167
 DB 129 PDVWAGAGTVTVSSSEPARFTIYPLFPQALSSDVIIGCLIHDFPSGIMVMTWGSXG 188
 QY 168 TILSVSL--ELQDSGTWCTCTVLQNKQVEFKIDYPCPAPEKSCDKTH-----T 215
 DB 189 DITTVNFPALASGRYTWSS-----SOLTLPAVECEGESVKCSVOHDSNPVELNVN 241
 QY 216 CPELLGSPVLPFPKPK-----DTLMSRIPEVYCVVVDVSHDEP-VKFNW 262
 DB 242 CEGTCSPTT---PPPSQPSLSLQRPALIEDLIGSDASTCTLNGL-RDPGAVITW 296
 QY 263 YVDGVEVNAKTKPREBOYNST---YRVSVLTVLHODMLNGEKYCKVSNKALPAPTEK 319
 DB 297 -----EPSTGKQAVQKAVQNSCGCYSVSVLPGCAERNNSGASKCTVYTHESDT-LTG 350
 QY 320 TTSKAKGQPREPOVYTLTPSRDELTKNQ-VSLTCLVKGFPYSDIAYEBSNQ---PENN 375
 DB 351 TIAKTVTVTFPQVHLPPSEELINELVSLTCLVRAFNPKYVLRVLMHGNEELSPSY 410
 QY 376 YKTPPEVLDSG--SFPLYSKLTVDKSRMQGNVSCSMHMLNHYTOKSLSPG 431
 DB 411 LVFEPLKEPGBGATTYLVTSVLRSALTIWKQGDQYSCVGHGHALPMFTOKTIDRLSG 468

RESULT 71
 Q96KX8 PRELIMINARY; PRT; 496 AA.
 AC Q96KX8;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strauberg R.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC016369; AAH16369.1; -
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 496 AA; 53391 MW; D346929849040D69 CRC64;

Query Match 12.3%; Score 297.5; DB 4; Length 496;
 Best Local Similarity 25.8%; Pred. No. 2,4e-16;
 Matches 131; Conservative 60; Mismatches 226; Indels 91; Gaps 22;

QY 8 RHLLLVQLALP-----AATQGNKTVLGKGGTVLELTCTAS-----QKSIQFMKN 57
 DB 2 KHLWFFLLVAAPRWVLSQLQDESGLVKSSTSLTLCTVSGSISSSSYWGWLRQP 61
 QY 58 QIKLIGNQSFLLTGPSPKLNDRADRSRL---WDQGNFPLIKNLKIEDSDTYICEVEDQ 114
 DB 62 PGKLEMIANTYYSGITVYVPSLKRITISVDTSKNDLSLKRVSATADTVYCAHGY 121
 QY 115 KEEVQLLVPLGFLANSDFHLLQGGSLTLTLESPGSSPSV-----QCRSPRGKN-----IQ 164
 DB 122 SRS-----GRTGAIID---YWGQGLTVTSASPTSPKVPPLSLCTQPDGNVVIACLVQ 172
 QY 165 G---GKTLVSQLE-----LQDSG-----TWCTVYLQNKQVEFKI-- 197
 DB 173 GFPPQEPVLSVTSWSSGGCVTARNFPPSQDASGDLVTTSSQTLTPATQCLAGKSVTCVKA 232
 QY 198 -----DI-VPCPAPEKSCDKTHCPCLLGG---PSVFLPFPKPDYLMISRTPEYTCV 247
 DB 233 YTNPSQDVTPCPVSPPTPSPSTPTTPSPSCCHPLSLHRLPDL-LLGSANLTCT 291
 QY 248 VVDVSHDEPEYKFNWYVDGVEVNAKTKPREBOYNSTYRVSVLTVLHODMLNGEKYCK 307
 DB 292 LTGL-RDASGVTFTWTPSSGK--SAVQGPBERDLGCGYSSSVLPGCAERNNHGKFTCT 348
 QY 308 VSNKALPAPTEKTSKAKGQPREPOVYTLTPSRDELTKNQ-VSLTCLVKGFPYSDIAYE 366
 DB 349 AAYESKTPPLATLSKS-GATFPREVHLPPSEELINELVLTCLARGLSPQDVLRW 407
 QY 367 ESNQ--PENNYKTPPVL-DSD--SFPLYSKLTVDKSRMQGNVSCSMHMLNHY 421
 DB 408 LQSGQELPREKYLTMARQSPSGTTFPAVTSILRVAEDMKKGDYSCVGHGHALPLAF 467
 QY 422 TOKSLSLSPG-----LQDETC 438
 DB 468 TOKTIDRLAGKPTHVNVSVMAAEVDGTC 495

RESULT 72
 Q8NCL6 PRELIMINARY; PRT; 493 AA.
 AC Q8NCL6;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical protein FLJ90170.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Isegai T., Oca T., Nishikawa T., Hayaehi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
 RA Masuno Y., Oho T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
 RL "NEDO human cDNA sequencing project."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK074651; BAC1114.1; -
 DR InterPro; IPR003599; Ig_
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guetlinck S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaez P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Sessa T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilmink L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohenki S.,
 RA Hayashizaki Y.,
 RT "functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK002875; BAB2422.1; -
 DR PIR; F33932; F33932.
 DR HSSP; P01810; 2FBJ.
 DR InterPro; IPR007110; Ig_1Lke.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00407; Ig1; 2.
 DR PROSITE; PS50835; IG_LIKE; 2.
 DR PROSITE; PS00290; IG_MHC; 2.
 SQ SEQUENCE 426 AA; 45819 MW; 56E1275BA48F6FB1 CRC64;

Query Match 11.9%; Score 286.5; DB 11; Length 426;
 Best Local Similarity 26.8%; Pred. No. 1.6e-15;
 Matches 106; Conservative 58; Mismatches 140; Indels 91; Gaps 18;

QY 96 IKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLGGOSLTLTLESP-----PGS 149
 DB 45 LSLTSEDSAVVFCARSDYGYFAMDYWG-----QGTSVTVSSSARNPTIYPLT 94
 QY 150 SPVQGRSP-----RGNQ-----GGKTLVSQGL-- 174
 DB 95 LPRALSSDPYIIGCLIHDPFPGSGTMVYWGSGKIDITVNFPPALASGGGYTWS-SQLTL 153
 QY 175 ---ELDQSGTWCTVYLVONQKVEFKIDI-----VCPAPRPPKSCDKTTCPELGGPSV 225
 DB 154 PAVECEGESVYKCSVGHDSNAVQ-ELDVKCSGPPPCP-PCPPSCG-----PSL 200
 QY 226 FLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFMYVYDGVENHAKTKPREQYNT- 284
 DB 201 SLQRPALPD-LLLGSDASITCTLNGLRNPEGAV-FTW-----EPSTGKDAVQKAVQNSC 253
 QY 285 --YRVSVTLVHODMLNGKEYCKVSNKALPAPIKTIKSKAKGQREQVYTLPPSRBE 342
 DB 254 GCYSVSVLPGCAERNNSGASFCVTHPESDT-LGTIAKTIVTVPQVHLLPPSBE 312
 QY 343 LTKNQ-VSLTCLVKGFPSPDIAVEMESNGQ--PENNYKTTTPVLDSG--SFFLYSKLT 396
 DB 313 LALNELVSLTCLVRAFPKEVLYRWLHGNBELSPESYLVFPELKEPGEATTYLVTSVLR 372
 QY 397 VDKSRWQGNVPSGSVMHEALHNHYTKSLSPG 431
 DB 373 VSAELWKQGDQYSCVNGHEALPMNFTQKTIIDRLSG 407

RESULT 77
 Q91WP5 PRELIMINARY; PRT; 479 AA.
 AC Q91WP5; 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC013656; AAH13656.1; -
 DR InterPro; IPR007110; Ig_1Lke.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SM00406; IgV; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 479 AA; 51603 MW; ECH2D0877748584F CRC64;

Query Match 11.9%; Score 286.5; DB 11; Length 479;
 Best Local Similarity 24.0%; Pred. No. 1.9e-15;
 Matches 126; Conservative 80; Mismatches 160; Indels 159; Gaps 29;

QY 1 MNRGVPRHLVLVIGLALLPAPATQ---GNKVLAKKGDVETLTCTAS----- 44
 DB 1 MNFGLTL--IFVLTLKGVQCEVQLVESGGGLV-KPGSLKVCASAGLFRSNYAMSV 56
 QY 45 -QKSLIQFHWK---NSNQIKILGN-----QGSFLTKGPKSLNDRADSRSLMDQGNP 93
 DB 57 RQSPKRLRWVAALNSNG---GNTYVSDTMKGRFTI-----SRDAKSTLY----- 99
 QY 94 LIIKNLKIEDSDTYICE-----EVEDQKE-EVQLVFGLTANSDTHL- 134
 DB 100 LQMSLRSEDTAFYICVAGGYFDVWGAGTAIVYSSEPARPEPTIYTLTPQALSSDPYIIG 159
 QY 135 -----QGOSLTLTLESPG-----SSPSVQGRSPRGN 163
 DB 160 CLIHDPFPGSGTMVYWGSGKIDITVNFPPALASGGRYTWSQLTLPAVEE--DEGESV 216
 QY 164 QGKTLVSQGLDELQSGTWCTVYLVONQKVEFKIDIVCP-----APRPPKSCDKTTC 216
 DB 217 K-----CSVQHDSNPVQ-ELN-VNCPGICSPPTTPPPSC----- 250
 QY 217 PELLGSPVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPE-VKFNWYVDGVENHAKTK 275
 DB 251 ---PSLSLQRPALPD-LLLGSDASITCTLNGLRNPEGAV-FTW-----EPSTGKDA 296
 QY 276 PREQYNT--YRVSVTLVHODMLNGKEYCKVSNKALPAPIKTIKSKAKGQREQV 332
 DB 297 VQKKAQVNSCGYSVSVLPGCAERNNSGASFCVTHPESDT-LGTIAKTIVTVPQVHLLPPSBE 355
 QY 333 VTLTPPSDELTKNQ-VSLTCLVKGFPSPDIAVEMESNGQ--PENNYKTTTPVLDSG- 387
 DB 356 VHLIPPEBELALNELVSLTCLVRAFPKEVLYRWLHGNBELSPESYLVFPELKEPGE 415
 QY 388 -SFFLYSKLTVDKSRWQGNVPSGSVMHEALHNHYTKSLSPG 431
 DB 416 TTYLVTSVLRVSAELWKQGDQYSCVNGHEALPMNFTQKTIIDRLSG 460

RESULT 78
 Q7TMK4 PRELIMINARY; PRT; 479 AA.
 AC Q7TMK4; 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marzella K., Farmer A.A., Rubin G.M., Hong L.,

Db 6 YLTTATPGLSLVQOEGSPGLVKRSQSLSLTCTGYSITSIGYMMWMTIRQP--GNKLE 65

Oy 66 --GSFLTKGPSKLNRADSRSL---WDQGFLLIKNLIKEDSDTYIC----- 109

Db 66 WMGYINYGSSNNYNPSLNKRISITRDTSKNGFFLKNSVTEDTATYYCASRGYSWPNM 125

Oy 110 -----EVEQKEVQLVRFGLT---ANSDFHLL-----OQOS 138

Db 126 GGGTLVTVASBASRPNT---IYPLTLPPALSSDPVIIGCLIHDPSPSGTMVYMGKSGKD 182

Oy 139 LTLTLESPG-----SSPYOCRSFRGNIQGKTLVSQLELDGSGTWICT 185

Db 183 IT-TNFPFAPALASGRRTYMSQULTIPAYNC--PEESVK-----CS 220

Oy 186 VLQNQKKVEFKDIDIVPCPAPEBKSCDKHTHCPELLGSPSVLPFKPKDTIMISRTPEVT 245

Db 221 VQHDSNPVQ-ELD-VNCGSPTPRPTITIPSCQ-----PSLSQRLALBD-LILGSDASIT 272

Oy 246 CVVDVSHDEPDYKRNMYVDGVENHNATKPREEQYST--YRVVSVLTVLHDQWLNGK 302

Db 273 CTLNLRNPBEGAV-FTW-----EPSTGDVAQKKAONS CGCYSVSYLPGCAERMSGA 326

Oy 303 EYKCVCVSKALPAPIEKTIISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGYFSD 361

Db 327 SFKCVTHPE-SGITLTGIATAKTNTTPFPQVHLPPSEELANLNLSTICLVRAFNFKE 385

Oy 362 IAVEMESNGQ--PENNYKTPPVLDSDG--SFPLYSLTVDKSRMOOGNVFSCVMHEA 416

Db 386 VLVRLLHGNEBELSPESLYLFEBLKBGEATTYLTVSVLRVSAETWKQGDGYSCMVGHFA 445

Oy 417 LHNYTKXSLSLSPG 431

Db 446 LPWNFTOKTIDLRLSG 460

RESULT 83

Q29028 PRELIMINARY; PRT; 99 AA.

ID Q29028 AC Q29028;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE CD4, allele 2 (Fragment).

OS Sus scrofa (Pig).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.

RN NCBI_Taxid=96823;

RX [1]

RP SEQUENCE FROM N.A.
RC STRAIN=#1183; TISSUE=Blood;
RX MEDLINE=93329116; PubMed=8335933;
RA Gustafsson K., Germana S., Sundt T.M., Sachse D.H., Leguern C.;
RT "Extensive allelic polymorphism in an exposed region of the
miniature, J. Immunol. 151:1365-1370(1993).
RL EMBL; X65630; CAA46584.1; -.
DR PIR; I47132; S21462.
DR HSSP; P01730; ICDD.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGv_1.
DR PROSITE; PS00835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 99 AA; 11390 MW; C22955BEA22831BF CRC64;

Query Match 11.6%; Score 280.5; DB 6; Length 99;
Best Local Similarity 56.1%; Pred. No. 6,8e-16;
Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

32 KKGDTEVELCTGASQKKSIOFHMKNSNQIKILGNQGSFLTQGP-SKUNRADSRSLWDQG 90

Db	1	YAGDLAEIPCHSSQKKNLPBFWMKNSDQIKILRSRNLHMKASVTELSRLDSKKNMDDHG	60
Qy	91	NEPLIIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTAN	128
Db	61	SFPLIIKNLEVTDSGIVCEVEDKRIEVOQLLVFRLTAS	98
RESULT 84			
Q9BRV0		PRELIMINARY;	PRT; 500 AA.
ID	Q9BRV0		
AC	Q9BRV0;		
DT	01-JUN-2001 (Tremblrel. 17, Created)		
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)		
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)		
DE	Hypothetical protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Prostate;		
RA	Strasbourg R.;		
RL	Submitted Apr-2001 to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC005951; AAH05951.1; -.		
DR	HSSP; P01789; IMCP.		
DR	InterPro; IPR007110; IG_1like.		
DR	InterPro; IPR003006; IG_MHC.		
DR	InterPro; IPR003596; IG_v.		
DR	Pfam; PF00047; Ig_4.		
DR	SMART; SM00406; IGV_1.		
DR	PROSITE; PS50835; IG_LIKE; 4.		
DR	PROSITE; PS00290; IG_MHC; 1.		
KW	Hypothetical protein.		
SQ	SEQUENCE 500 AA; 54154 MW; 0A9BF43F2A3CC6D9 CRC64;		
Query Match 11.3%; Score 273.5; DB 4; Length 500;			
Best Local Similarity 25.8%; Pred. No. 2.4e-14;			
Matches 104; Conservative 56; Mismatches 154; Indels 89; Gaps 19;			
Qy	86	LMDGN-----PPLIIKNLKIEDSDTYICEVEDQKEEVOLLV-----GL	125
Db	136	VWGKGTIVTSSASPTSPKVFPLSLCSTQPDGNVVIACLVGGFPPEPLSVTWSSEGGV	195
Qy	126	T-----ANSDTHLQGSILTLTLESPPGSSPSVQCRSPRGKNIQGKTLVSQLEL	176
Db	196	TARNPPSQDASGDLVTTSSQ---LTL-----PARGCLA--GKSV-----	230
Qy	177	QDSGTWCTVLQONQKVEFKIDI-VPCPAPRPKSCDKHTTCPELLIG--DSVFLPPKP	232
Db	221	-----THVHGKYNTPSQ--DVTVPCEVPSTPTPPSPSTPTPSPSCCHRLSLHRL	281
Qy	223	KDPLMISTPEVTCVVDVSHDEBVKWVVDGVEVNMATKPREBQYNSTYRVSVLT	292
Db	282	ED-LILGSEANLTLTGL-SDASGVFTWPPSSGK--SAVQGPDRDLCCGYSSVLS	337
Qy	293	VHADQMLGKYEKKVCNSKALPAPLEKTIISAKAQPREPOVYTLPPSRDELTKAQ-VSLT	351
Db	338	GCAEPMNHGKFTCTAAYPESKTPTLTLSKS-GTFPPEVHLHPPESEELALNELVLT	396
Qy	352	CLVKGFPSPDIAVWESNGQ--PENNYKTPTPVD-SDG--SFLYSKLTVDKSRMOGN	406
Db	397	CLAGFPSPKVDLVWMLQSGQELPREKVLTVASRQEPSSGTTTFAVTSTLRVAEDMKGD	456
Qy	407	VFSCSVMEALHNHYTKSLSLSG-----LQDLETG	438
Db	457	TFSCVNGHEALPLAFTOETIDRLAGKPTHVNVSVVMAEVDGTC	499
RESULT 85			
Q8K0Z4		PRELIMINARY;	PRT; 480 AA.
ID	Q8K0Z4		
AC	Q8K0Z4;		

```
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to expressed sequence A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_Taxid=10090;
RP [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: BC029188; AAH29188.1; -.
DR InterPro: IPR003599; IG_1like.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003596; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; IG_4.
DR SMART: SM00409; IG_3.
DR SMART: SM00407; IGc1; 3.
DR SMART: SM00406; IGc1; 1.
DR PROSITE: PSS0835; IG_LIKE; 4.
DR PROSITE: PSS0290; IG_MHC; 2.
SQ SEQUENCE 480 AA; 51645 MW; 8690A63C69CDBED CRC64;

Query Match 11.2%; Score 270; DB 11; Length 480;
Best Local Similarity 24.7%; Pred. No. 4.5e-14;
Matches 116; Conservative 71; Mismatches 178; Indels 104; Gaps 22;

QY 30 LGKKGDTVELTCTAS--QKSIQFHMKNSQIKILNQG---SFLT-KGSGKLNDRAD 81
DB 30 LKPKGLVAVKISCASQYTTSPDISW---MKQPGQGPBEMIGWISPDGSSSEIVEKK 84
QY 82 SRRSLW---DOGNFPLIKLKIEDSDTYICEVEDQKEVQLVFGLTANSD---THLL 134
DB 85 GKATLVADKSSNMTAVMHLSSLSSENSAVFC-----ARSLKGAFAYWG 127
QY 135 QGOSLTLTLES-----PGSS-----PSVQCRSPRGKN----- 162
DB 128 QGTLTVSAESAANPTIPLPLPALSDPVIIGLIHDYFPGTWNVTWKGSKGDKITTV 187
QY 163 -----IOGKTLVSQI-----ELDQSGTWCTVLQNGKYEFKIDIPCAPEKSCD 211
DB 188 NRPPLASGGRYTMSSQTLTPAVECPGESSVKCSVOHDSNPVO-ELD-VNCSGPPPPPI 245
QY 212 KTHCPPELLGSPVFLPPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGEVHN 271
DB 246 TTPSCQ-----PSLSLQRPALBD-LLLGSDASITCTLNGLRNPEGAV-FTW-----EPST 293
QY 272 AKTPREEDYNST--YRVSVTLVHQMNLNGKEYKCVSKNKALEPIEKTISKAKGP 328
DB 294 GDAVQKKAIVONSCGCVSVSVLPGCAERWNSGASFKCTVTHE-SGTLTGITAKVTNT 352
QY 329 RRPQVYTLPPSRDELTKNQ-VSLTCLYKGFYPSDIAVESNQ---PENNYKTPPVLD 384
DB 353 PFPQVHLLPPSEELALNELSLTCLVRAFNPEKEVILVRHLGNEELSPSYLVFEFLKP 412
QY 385 SDG--SFLYSKLTVDKSRMOQGNVSCVMHEALNHNHTOKSLSLSPG 431
DB 413 GEGATTYLVTSVLRVAEATWKGQDQYSCVNGHEALPMNTQKTIIDLSG 461

RESULT 86
Q90MB5 PRELIMINARY; PRT; 482 AA.
AC 090MB5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor.
OS Anas platyrhynchos (Domestic duck).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
RN NCBI_Taxid=8839;
RP [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Chan S.W.S., Middleton D.L., Lundqvist M., Watt G.W., Higgins D.A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF378701; AAK59279.1; -.
DR InterPro: IPR003599; IG_1.
DR InterPro: IPR007110; IG_1like.
DR Pfam: PF00047; IG_4.
DR SMART: SM00409; IG_3.
DR PROSITE: PSS0835; IG_LIKE; 2.
KM SIGNAL.
FT SIGNAL.
SQ SEQUENCE 482 AA; 54624 MW; BDA0B8BC6172AD0B CRC64;

Query Match 7.8%; Score 189; DB 13; Length 482;
Best Local Similarity 22.2%; Pred. No. 2.8e-07;
Matches 95; Conservative 70; Mismatches 148; Indels 114; Gaps 22;

QY 8 RLHLVLQALALPAATQGNKVVLGKKGDTVELTCTASQKSIQFHK-----NSNQIK 60
DB 12 RAVFVLQGLL-THIMHQQIGVEKEVILNC---KXHDQVWTKVEYDAGSSAIIIQ 66
QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDQKEVQL 120
DB 67 ILAGK---IFKRAVMSRSET---NQNSHGLKVNLSRIDAGTYICEGSDHNSISL 118
QY 121 LVFGLTANSPTHLQGSITLT-LSPGSSPS-----VQCRSPRGKN 162
DB 119 HVKLTISSNGYFLGDDIELTWMKSPKQPREITLPHNSNRYTPEVLQNETPQ--- 175
QY 163 IOGKTLVSQIQLDQSGTWCTVLQN---QKVEFKIDIVPCAPPEKSCDKTH 214
DB 176 ---KVALVKQLQPPDSGWTICMHSDSPSINENISFNVKVL-----GPEKHLERM 224
QY 215 -----TCP-----ELIGPSVFLPPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFN 261
DB 225 YAAVDSTVLSWHLNFRKIGWKEFP-----TGQINMOEGNAITYELD-----FN 269
QY 262 WYVDGEVHNKATKR-----EEQYNSYRVSVTLVHQMNL-NGKEYCKV--SNKA 312
DB 270 ATADG-ELRETKKSQALLLEPEMKRDSITVEY---KIKIKQLHSGEYTCQLYNNRY 323
QY 313 LPAPIEKTISKAKGPREPQVYTLPPSRDELTKQVSLTCLYKGFYPSDIAVES-NGQ 371
DB 324 IQSKTELVMQVSNRPPR---LPKA-----EMTLLQVSSPIPNVHLWERNVGT 373
QY 372 PENNYKT 378
DB 374 KMDGKKS 380

RESULT 87
Q90524 PRELIMINARY; PRT; 268 AA.
AC 090524;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Novel antigen receptor (Fragment).
OS Ginglymostoma citratum (Nurse shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Orectolobiformes;
OC Ginglymostomidae; Ginglymostoma.
RN NCBI_Taxid=7801;
RP [1]
RP SEQUENCE OF 1-240 FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=95183140; PubMed=7877689;
RA Greenberg A.S., Avila D., Hughes M., Hughes A., McKinney E.C.,
```

RA Flajnik M.F.;
 RT "A new antigen receptor gene family that undergoes rearrangement and
 RT extensive somatic diversification in sharks."; [NCBI](#)
 RT Nature 374:168-173 (1995).
 RU [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RA Greenberg A.S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U18680; AAB48352.1; -
 DR HSSP: P01842; 7FAB.
 DR GO: GO:004872; F:receptor activity; IEA.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig: 2.
 DR SMART: SM00407; Igcl. 1.
 DR PROSITE: PS50835; IG_LIKE; 2.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Receptor.
 FT NON TER 268 268
 SQ SEQUENCE 268 AA; 29579 MW; 0CDE524DDE9FDC4 CRC64;

Query Match 7.7%; Score 186; DB 13; Length 268;
 Best Local Similarity 24.3%; Pred. No. 2.2e-07;
 Matches 65; Conservative 39; Mismatches 98; Indels 66; Gaps 10;

QY 223 PSVFLPFPKPDLMISRTPE-----VTCVVVDVSHDEPKFNNYVVG----- 266
 DB 14 PIVF-----TARVDTPDEITETGEBLSINCVLDSNCALSTYWNKRGSGSTNEE 65
 QY 267 -----VEVNAKTKPREEQYNSTYRVSVLTALHDMWNGKEYCKV--SNKALPAP 316
 DB 66 TISKGRVETVETNSGSKSFSLRIND-----LTV-----EDSGYRKQVYRKMAVYDGC 113
 QY 317 IEKT-----ISRAKGPREFPYVTL--PPSRDELTKNOVSLTCLVKGFPYSDIA 363
 DB 114 LEELDMIVYVGCGTGVTPNPGIPLSPPIVSLHSAITEQRANGFVQLCLISGYYPENIA 173
 QY 364 VEWESNGQP-ENNYKTPTPVLDSDGSFPLYSKLTVDKSRMOQGNFSGSVNHEALHNYT 422
 DB 174 VSMQNTKTTITTSGFATTSPTVKTSNDFSCASLTKVLPQWMSKVSVCQVSHSATYSNQR 233
 QY 423 OKSLSLSPQLQDETCAEADQDELGLW 450
 DB 234 KEIRSTS-----ELAVILRPDPTVEIWL 255

RESULT 88

Q9DBM4 PRELIMINARY; PRT; 130 AA.
 AC Q9DBM4;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE 1810027001rik protein.
 GN 1810027001rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Pancreas;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Aochi J., Fukuda S.,
 RA Saito T., Okazaki H., Gojohori T., Bono H., Kaubawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischnann W., Gaasterland T., Glasl C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guenich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombauer P.,
 RA Nordone P., Ring B., Rungwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
 RA Wyszynski B., Yoshida K., Haegawa Y., Kawaji H., Kohzuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."; [NCBI](#)
 RL Nature 409:685-690 (2001).
 DR EMBL: AK007622; BAB5142.1; -
 DR PIR: S22760; S22760.
 DR HSSP: P01842; 7FAB.
 DR MED: MGI:1916425; 1810027001rik.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig: 1.
 DR SMART: SM00407; Igcl. 1.
 DR PROSITE: PS50835; IG_LIKE; 1.
 DR PROSITE: PS00290; IG_MHC; 1.
 SQ SEQUENCE 130 AA; 14253 MW; 438197975E766E54 CRC64;

Query Match 7.6%; Score 183.5; DB 11; Length 130;
 Best Local Similarity 36.5%; Pred. No. 1.3e-07;
 Matches 38; Conservative 18; Mismatches 43; Indels 5; Gaps 3;

QY 326 GQPR-EPQVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVWESNGQP-ENNYKTPPVLT 383
 DB 25 GQPKSSPSVTLFPSSSELETNKATVCTITDFPGVGVTDWKVDGTPVQGMETTPSK 84
 QY 384 DSDGSFPLYSKLTVDKSRMOQGNFSGSVNHEALHNYTKSIS 427
 DB 85 QSNKNYMASSVTLTPADAMERHSSVCQVTHE--GTVKSLIS 125

RESULT 89

Q8N355 PRELIMINARY; PRT; 234 AA.
 AC Q8N355;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Strauberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC028090; AAB28090.1; -
 DR PIR: S12441; S12441.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: PF00047; Ig: 2.
 DR SMART: SM00407; Ig: 2.
 DR SMART: SM00406; Igcl. 1.
 DR SMART: SM00406; Igcl. 1.
 DR PROSITE: PS50835; IG_LIKE; 2.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Hypothetical protein
 SQ SEQUENCE 234 AA; 24792 MW; CC848CABBA4A9D53 CRC64;

Query Match 7.6%; Score 183.5; DB 4; Length 234;
 Best Local Similarity 25.0%; Pred. No. 3e-07;

Matches 55; Conservative 39; Mismatches 99; Indels 27; Gaps 7;

QY 210 CKHTHCPELLGSPSVFLPRPKKDTLMSRTEPVTCVVAVDVSHDEPVKFNMY----- 263
DB 14 CGSVTSYVLTQPPSVVAP-----GQTARITCGGNNIGSK---SVHYQQKPFQ 60
QY 264 VGVGVHNAKTKPR---EEQYNSTYRVSVLTVLHODMNLNGKCKVKNKALPAPI--- 317
DB 61 AVLVVVYDSDRSRSGIPERFSSGNSGNTATLTISRADADDEADYTCQLWSSSDHVFVG 120
QY 318 EKTISAKAQPR-EPQVYTLPRSRDELTKNQVSLTCLVKGFYPSDIAVWESNQP-ENN 375
DB 121 GGGTKLTVLQCPKAPAVTTLFPPSESELOANKATLVCLISDFYGAVTAMKADSPVKAG 180
QY 376 YKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFCSCVME 415
DB 181 VETTPSKQNNKMYASSYLTLTAKAMETHSSYSCQVTHE 220

RESULT 90
Q99M11 PRELIMINARY; PRT; 235 AA.
AC Q99M11
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RL Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002129; AAH02129.1; -
DR HSSP; P01703; 7FAB.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 2.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS00835; IG_Like; 2.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 235 AA; 25403 MW; 39807BE6782A3FB CRC64;

Query Match 7.6%; Score 183.5; DB 11; Length 235;
Best Local Similarity 24.9%; Pred. No. 3e-07;
Matches 58; Conservative 39; Mismatches 101; Indels 35; Gaps 9;

QY 213 HTTCPELLGSPSVFLPRPKKDTLMSRTEPVTCVVAVDVSHDEPVKFNMY----- 263
DB 15 TSSCAQLV-----LTQPSVSVF-SLGSYAKLPCKASTGNIGDSYV--NMVQOYMGSRPT 65
QY 264 --VDGVEVNAKTKPR---EEQYNSTYRVSVLTVLHODMNLNGKCKVKNKALPAP 316
DB 66 NMIVGDLPKSGVSDPSSGIDSSNSAP-----LTQNVQADDEADYTCQSYSGIRVF 120
QY 317 IERTISKAKQPR-EPQVYTLPRSRDELTKNQVSLTCLVKGFYPSDIAVWESNQP-EN 374
DB 121 GGGTKLTVLQCPKAPAVTTLFPPSESELETKATLVCTISDFYGVVTVDMKADGTPVVG 180
QY 375 NKTTTPPVLDSDGSFFLYSKLTVDKSRWQGNVFCSCVMEALHNHYTKSL 427
DB 181 GVEITQPSKQNNKMYASSYLTLTAKAMETHSSYSCQVTHE--GHTVEKSL 230

RESULT 91
Q723Y4 PRELIMINARY; PRT; 236 AA.
AC Q723Y4
DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.U., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzyminski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RC Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005332; AAH05332.1; -
KW Hypothetical protein.
SQ SEQUENCE 236 AA; 25702 MW; 7E9F8AED23084BC6 CRC64;

Query Match 7.5%; Score 182; DB 4; Length 236;
Best Local Similarity 23.1%; Pred. No. 4e-07;
Matches 75; Conservative 36; Mismatches 107; Indels 106; Gaps 10;

QY 11 LVLVLDLALPAA-----TQGNKVVLGKKDPTVELTCTASQKSLQIFHWKNSNQIKLGN 64
DB 9 LIGLLILGFPGARCDIOWTQSPSSLASVGDVITTCRASQDISNYLWFFQKP-----GK 64
QY 65 QGSFLLTKGPKLNDRAISRSLMDQG-NPPLITKLTKEPDTYCEVDQKEVQLLVF 123
DB 65 AKPSLIYGASSLQSGVQSGFSGSGGTPRTLTISLQEDPRTYTCQ----- 111
QY 124 GLTANSDTHLQGGSLTTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSOLELQDSGTWT 183
DB 112 -----QYKSVPTVF-----GQGT-- 124
QY 184 CTVLNQKKKVEPKIDIVCPAPPEKSCDKHTCPPELLGSPSVFLPRPKKDTLMSRTE 243
DB 125 -----KLEIK-----RTVAPSVFIRP--PSDEQLKSGTAS 153
QY 244 VTCVVAVDVSHDEPVKFNMYVGV--EVHNAKTKREBOYNSTYRVSVLTVLHODMNLNG 301
DB 154 VVCLNNYTPR--EAKVQKVDNALQSGNSQESVTEQDSKDSITYSLSTLTLSKADYEKH 211
QY 302 KEKCKVSNKALPAPIEKTISKAK 325
DB 212 KYVACEVTHHGSLSPVTKSFNRGE 235

RESULT 92
Q8NFP4 PRELIMINARY; PRT; 955 AA.
AC Q8NFP4
DT 01-OCT-2003 (TrEMBLrel. 25, Created)

```

DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Glycosyl-phosphatidylinositol-MAM.
GN GPI-M.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2207705; PubMed=12082541;
RA De Juan C., Iniesta P., Gonzalez-Quevedo R., Moran A.,
RA Sanchez-Pernate A., Torres A.J., Balibrea J.L., Diaz-Rubio E.,
RA Cruces J., Benito M.,
RT "Genomic organization of a novel glycosylphosphatidylinositol MAM gene
RT expressed in human tissues and tumors.";
RL Oncogene 21:3089-3094(2002).
CC -1 SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF478693; AAM77220.1; -.
DR Genew; HGNC:19267; MDGAL.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016163; F:nitrogenase activity; IEA.
DR GO; GO:0009399; F:nitrogen fixation; IEA.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR000318; Nitrogense_compl.
DR Pfam; PF00047; Ig; 6.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00409; IG; 6.
DR SMART; SM00408; IGC2; 6.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS50835; IG_LIKE; 6.
DR PROSITE; PS50060; MAM 2; 1.
DR PROSITE; PS00699; NITROGENASE 1; 1.
KM Glycoprotein: Immunoglobulin domain.
SQ SEQUENCE 955 AA; 105790 MW; BD41A1EB10A05962 CRC64;

Query Match 7.5%; Score 181.5; DB 4; Length 955;
Best Local Similarity 21.8%; Pred. No. 3e-06;
Matches 99; Conservative 55; Mismatches 151; Indels 149; Gaps 18;

QY 36 TVELTCTASQKKSIOFHKNSNOIKILNGSGFLTKGSKLANDRADSRSLMDQGNFPLI 95
DB 152 TVFLKCTVSNPNPAPFIMKRGSD-----TLSSHODNG-VDIYEPLYTOGETKVL 199
QY 96 -IKNLKIEDSDTYICEVEDQKE---EVLVLVGLTANSDTHLQGSGLTLTLESPPGSSP 151
DB 200 KLNKLRPODYASTYQGVSVANVCGLIPDKAITRLNTTAPRLK-LISNVELLVANPGSNV 258
QY 152 SVQC-----RSPRGKNIQGGKTLVSQLELQDSGTWCTVLON--- 189
DB 259 TVQCCLLTGGDPLPQLQWSHGPGRLPLGALAQGG-TLSIPVQARDSGYVNCATANNVNP 317
QY 190 -----QKKEFKIDIVPCAPAPBKSCDKHTHCPBELLGSGSVLFPKPKDTLMI 238
DB 318 AKKTVNLVRSKNKATFQI-----TPDV-----KESENQL 349
QY 239 SRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHDM 298
DB 350 GQDLTLKSGVDVAPPE--KVTYQWFKNGKPARMSKRLVTRNDPLPAVTSLELIDHF 407
QY 299 LNKKEYKCKVSNKALPAP--LEKTIKAKQGPPEPVYTLPPSRDELTKQVS---LTC 352
DB 408 SDYGYTLGWASPPGAPVPLDLEVENIS---SETVPRTISVKGAAVTVREGSFALQC 463
QY 353 LVKGYPSDIAVEMESNQGPENNYKTTTPVLDSDGSFLYSKLTVDKSRMQGNVFSQSV 412
DB 464 EVRG-----KRPPEVLMS-----RVDK----- 480

```

```

QY 413 MHEALHNHYTQSLSPGLQDETCAEADGEL 446
DB 481 -----EALLPSGLPLEBT-----PDGKL 499

RESULT 93
Q8BLX5 PRELIMINARY; PRT; 379 AA.
AC Q8BLX5.
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Weakly similar to signal-regulatory protein beta-1 precursor.
GN 9930027N05RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Aorta; and Vein;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK040982; BAC30768.1; -.
DR MGI; MGI:244824; 9930027N05RIK.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGc1; 2.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 3.
SQ SEQUENCE 379 AA; 42409 MW; 055BAE20759CF6D5 CRC64;

Query Match 7.5%; Score 181; DB 11; Length 379;
Best Local Similarity 20.1%; Pred. No. 9.3e-07;
Matches 88; Conservative 70; Mismatches 155; Indels 124; Gaps 17;

QY 5 VPFRHLVLVQLALPATOGNKVVLGK-----GDVELTCTASQKSI-----QFHWK 54
DB 10 IPHCVLILLILLGKGAVALKLVQPVKSPFVGAGSATLNTCVYTLPLPGPIKWRGV 69
QY 55 NSNOIKILNGSGFLTKGSKLANDRAD--SRSLMDQGNFPLIKNLKIEDSDTYICEVED 113
DB 70 GQSRLLIYPTGEYF---PRITSVDVKKRSNLD---FSIRISVTPADSGTYVC-VKF 121
QY 114 QKEVQLVLVGLTANSDTHLQGSGLTLTLESPPGSSPVQCSRRKNIQGGKTLVSQ 173
DB 122 QRG-----SEPDIEIGSGGTELSVAPK-SSPMWS--GPAANVP----- 160
QY 174 LELQDSGTWCTVYLQONOKVEFKIDIVPCAPAPBKSCDKHTHCPBELLGSGSVLFPKPK 233
DB 161 ---QQTVTFTC-----RSHG-----FFP----- 175
QY 234 DTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQYNSTYRVSVLT 293
DB 176 -----QNLTLKWFKNGLISHLTSEVERETSYSRVYSTVQV 213
QY 294 LHDQWLNGKEYKCKVSNKALP-APLEKTIKAKQGPPEPVYTLPPS-----RELTKNQ 347
DB 214 VLEPRDVSQITICVDVHTLDRAPL-----KEIAHISEIIVPTLLEISQPTWNV 266
QY 348 VSLTCLVKGYPSPDIAVEMESNQGPENNYKTTTPVLDSDGSFLYSKLTVDKSRMQGNV 407
DB 267 INVTCQIQKFYPRPFQVTLWLENGISRREVPFTHIVAKDGYNNISLVLVNISLLEBNMV 326
QY 408 FSCGVH-----EALHNH 420

```

DB 327 VTGCVHGDQAEVIETH 343

RESULT 94

Q8BFX8 PRELIMINARY; PRT; 397 AA.

ID Q8BFX8

AC Q8BFX8; 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Weekly similar to signal-regulatory protein beta-1 precursor.

GN 9930027N05RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;

RA [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Ovary, and Vagina;

RX MEDLINE=22354683; PubMed=12466851;

RA The PANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

DR EMBL; AK036935; BAC29645.1; -

DR EMBL; AK054545; BAC35818.1; -

DR MGD; MGI:2444824; 9930027N05RIK.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003597; Ig_c1.

DR InterPro; IPR003596; Ig_v.

DR Pfam; PF00047; Ig; 3.

DR SMART; SM00407; IGc1; 2.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG_LIKE; 3.

DR PROSITE; PS50835; IG_LIKE; 3.

SQ SEQUENCE 397 AA; 44157 MW; E1DE9DPAAC83D676 CRC64;

Query Match

Best Local Similarity 7.5%; Score 181; DB 11; Length 397;

Matches 88; Conservative 70; Mismatches 155; Indels 124; Gaps 17;

QY 5 VFRRLLLVQLALLPAATQGNKRVYLGKK-----GDTVELTCTASQKKS-----QFHWK 54

DB 10 IPHCVLLILLGLGAAVRELKVIQPVKSFVGAAGSATLNCVTYLLPVGPIMKMGV 69

QY 55 NSNQIKILGNQGSFLTKGSKLNDRAD-SRSLMDQGNFPLIIKULKIEDSTYICEVED 113

DB 70 GGSRLLIYPTGEYF-----PRITSVSDVKGRSLND---FSIRISVTPPADSGTYVC-VKF 121

QY 114 OKEEVQLLVFGLTANSDTHLQGSLLTLLESPPSSPSVQCRSPKRNKIQGKTLVSQ 173

DB 122 QG-----SSPEDEIQSGGTELSVFAPK-SSPEVVS--GPAARVP----- 160

QY 174 LELQDSGVTCTVLQNKVKVEFKIDIVPCPAPEPKSCDHTHCPELLGSPSVFLPPRPK 233

DB 161 ---QQTVPFTC-----RSHG-----FFP----- 175

QY 234 DTLMSRTPPEVTCVVVDVSHEDPEVKFMYVDGVEVHNAKTPRBEQVNSTYRVVSVLT 293

DB 176 -----QNLTLKMFKNGNESIHLETSSVEPESTSVSYRVSSTVQV 213

QY 294 LHQDWLNGKEYKCKVSNKALP-APLEKITSKAKGQPREPOVYTLPPS-----RDLTLTKQ 347

DB 214 VLEPDDVRSQITICEVDHTLDRPL-----RGLAHSEIIVQPTLETISQOPTMVMV 266

QY 348 VSLTCLVKGFPYSDIAVEWESNGQPENNYKTPPYLSDSGSFLYSKLTVDKSRMQGQV 407

DB 267 INVTCQIQKFYRRPQVTLNENGNSRREVPTTHLVNKGDTYNNWISMLLVNISALENNV 326

QY 408 FSCSYWH-----BALNH 420

DB 327 VTGCVHGDQAEVIETH 343

RESULT 95

Q8WU38 PRELIMINARY; PRT; 573 AA.

ID Q8WU38

AC Q8WU38; 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxId=9606;

RA [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA Strausberg R.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC021276; AAH21276.1; -

DR PIR; S21205; S21205.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR003596; Ig_v.

DR Pfam; PF00047; Ig; 4.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG_LIKE; 4.

DR PROSITE; PS00290; IG_MHC; 2.

KW Hypothetical protein.

SQ SEQUENCE 573 AA; 62967 MW; FD0722344033AC530 CRC64;

Query Match

Best Local Similarity 7.4%; Score 179.5; DB 4; Length 573;

Matches 124; Conservative 68; Mismatches 221; Indels 175; Gaps 26;

QY 1 MNRGVFRRLLLVLT--QALLPAAATQGNKRVYLGKKGDTVELTCTAS--QKSIQFHWKN 55

DB 1 MELGLSWIFLAILKGVCEVOLVESGGGLV---QGRSLRLSCAASGFTPDYAMHVR 57

QY 56 SNQIK-----ILGNQSF-----LTGPKSLNDRABRRBLMDQGNFPLIIKULKIDS 104

DB 58 QAPGKGLWVSGISWNSGISGVADSVKGRFTIS-RDAKNLSLYQMN-----SLRADT 110

QY 105 DTVICEVEDOKEEVQLVFGLTANSDTHLQGSLLTLLESPPSSPSV-----QCRSPR 159

DB 111 ALTYCAKSGSSSYI-----GYTGMD--VWGQGTIVVSSAPFKAPVPTIISGRAPK 162

QY 160 GKN-----IQGKTLVS-----QLELQDS----- 179

DB 163 DNSPVVLACLITGHPHSVTVTWYMGTSQIQRTPEIQRDSYYMTSSQLSTPLQWRQ 222

QY 180 GTWCTVYLQNKVKVEFKI-----DIVPCPAPEPKSC----- 210

DB 223 GEYKCVQHTASKSKKEIFRWPESPRAQASVPPAQAQPAEASLAKATPAATTNTRGCG 282

QY 211 -----DKTHCPELLGSPSVFLPPPKDQTLMTSRPEVTCVVVDVSH 254

DB 283 EEKKEKEKEKEQERETKTPPECPSHTQPLGYLLTPAVQD-LMIRDKATTFCFVNGSDLK 341

QY 255 DPEYKFMVYVDG-VEVHNAKTPRBEQVNSTYRVVSVLTVLHODWLNGKEYKCKVSNKAL 313

DB 342 DAHL--TWEVAGKVPTGVESEGLEBRHSNGSQSHRSLTLRLSLMNGTSVTCGLNHS 399

QY 314 P-----APLEKITS-KAKGQPREPOVYTLPPSDELTKNOVSLTCLVKGFP 359

DB 400 PPRQMLALREPAQAQPVYLSLNLASSDPPEAAEW-----LLCEVSGSP 444

QY 360 SDIAVEWESNGQPENN--YKTPPYLSDSGSFLYSKLTVDKSRMQGQVFSQSYWH- 415

DB 445 PNILMLLEDQREYNTSGFAPARPPQPGSTTFPAMSVLKVPAAPSPOPATYTCVSHED 504

QY 416 --ALHNHYTKSLSPGLQDETCAEADQELD-----GLWTT 452

DB 505 SRTLIN--ASRSLEVS-YLAWTPLIPQSKDENSDDYTTFDVSGLWTT 549

```

RESULT 96
096169 ID 096169 PRELIMINARY; PRT; 233 AA.
AC 096169;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007782; AA07782.1; -.
DR PIR; S12440; S12440.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004812; F:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR InterPro; IPR003006; IG_Like.
DR InterPro; IPR003596; IG_MHC.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE; PS50835; IG_Like; 2.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 233 AA; 24802 MW; C694F8397B276508 CRC64;

Query Match 7.3%; Score 176.5; DB 4; Length 233;
Best Local Similarity 26.5%; Pred. No. 1.1e-06;
Matches 50; Conservative 35; Mismatches 87; Indels 17; Gaps 6;

QY 240 RTPPTCVVNVSHDPEVKRWY-----VDGVEVNAHKIKR---EEQVNSTYRVSV 290
DB 35 QIARITCGSNLGSK---SVNWYQLRPGQAPILVYENKKEPACIPERLSATSEETAT 90
QY 291 LTVLQDWLNKKEYCKVSNKALPAIEKTISSKAK--GQPR-EPQVYTLPPSRDELTKNQ 347
DB 91 LTISVAVAGDEADYTCQVWDITTSQQYVFGTGTQVTVLQPKRANPTVTLPPSSSELOANK 150
QY 348 VSLTCLVKGFPSPDIWESNGQP-ENNYKTPPVLDSGSPFLYSKLTVDKSRMOQN 406
DB 151 ATLVLCLISDFYGAVTVAWKADGSPVAKGVETTKSKSNKAYASVLSLTPEQWKSJR 210
QY 407 VFSCSVNHE 415
DB 211 SYSCQVTHE 219

RESULT 97
090530 ID 090530 PRELIMINARY; PRT; 259 AA.
AC 090530;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Novel antigen receptor (Fragment).
OS Gingivomastoma citratum (Nurse shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Orectolobiformes;
OC OC NCBI_TaxID=7801;
RN [1]
RP SEQUENCE OF 1-235 FROM N.A.
RC TISSUE=Spleen;

```

```

RX MEDLINE=95183140; PubMed=7877689;
RA Greenberg A.S., Avila D., Hughes M., Hughes A., McKinney E.C.,
RA Flaig M.F.;
RT "A new antigen receptor gene family that undergoes rearrangement and
RT extensive somatic diversification in sharks."
RL Nature 374:168-173(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Greenberg A.S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U18687; AA848358.1; -.
DR HSSP; P01842; 7FAB.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG_Like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGV; 1.
DR PROSITE; PS50835; IG_Like; 2.
DR PROSITE; PS00290; IG_MHC; 1.
KW Receptor.
FT NON_TER 259
SQ SEQUENCE 259 AA; 28245 MW; 07F9860A92318B6E CRC64;

Query Match 7.3%; Score 176.5; DB 13; Length 259;
Best Local Similarity 24.2%; Pred. No. 1.3e-06;
Matches 64; Conservative 34; Mismatches 103; Indels 63; Gaps 11;

QY 223 PSVFLPPPKKPDITMSRTPE-----VTCVVDVSH-----ED 255
DB 14 PYVF-----TARVQTPRSVTKETGSESLITNCVLRDASVALGHTCWPFRKSGSTKE 65
QY 256 PEVFNWYVDGV-EVNAKTRPREEOYNSTYRVSVYTVLHOMLNKKEYCKVSNKALP 314
DB 66 NISPGGRYVETVNSVQNLINDLTVEDEGTTRCGSL-----YNN-----CSETIVLP 114
QY 315 -----APIEKTISSKAKQPREPOVYTL--PPSRDELTKQVSLTCLVKGFPSPDIWEME 367
DB 115 YAACGAGTAVTVN--PQIPSPPIVSLHSATEQGRANRFVQLVCLSGYPERIAVSMQ 172
QY 368 SNGQP-ENNYKTPPVLDSGSPFLYSKLTVDKSRMOQNVFSCVWHEALHNHYTQSL 426
DB 173 KNTKITITSGPATTSPVTSSNDFSCASILKVPLOEWSRGSVSCQVSHSATSSNORKEIR 232
QY 427 SLSPGLDDETCABQDGEIDGLM 450
DB 233 STS-----EIVALLRDPVTEEIM 250

RESULT 98
028260 ID 028260 PRELIMINARY; PRT; 739 AA.
AC 028260;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Vascular cell adhesion molecule-1.
OS Canis familiaris (dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Ballantyne C.M., Clubb F.J., Perrard J.L., Radovencovic B.,
RA Youker K.A., Smith C.W., Entman M.L., Hawkins H.K., Frazier O.H.,
RA Willerson J.T.;
RT "Increased Expression of VCAM-1 and ICAM-1 in Early Cardiac Allograft
RT Arteropathy in the Dog."
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U32086; AAA84866.1; -.
DR HSSP; P19320; IVCA.
DR GO; GO:0016020; C:membrane; IEA.

```

DR GO; GO:0016337; P:cell-cell adhesion; IEA.
DR InterPro; IPR003987; ICAM VCAM-1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig; 7.
DR PRINTS; PRO1472; ICAMVCAM1.
DR PRINTS; PRO1474; VCAM1.
DR SMART; SM00408; IGC2; 4.
DR PROSITE; PSS0835; IG_LIKE; 5.
DR Immunoglobulin domain.
KW Immune 739 AA; 81412 MW; BB5DA3853739C615 CRC64;

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC032452; AA32452.1; -
DR PIR; S12441; S12441.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 2.
DR SMART; SM00409; IG; 2.
DR SMART; SM00407; IGcl; 1.
DR SMART; SM00406; IGv; 1.
DR PROSITE; PS50835; IG_LIKE; 2.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 233 AA; 24961 MW; F092CFB6AA6E3A9A CRC64;

Query Match 7.2%; Score 174; DB 4; Length 233;
Best Local Similarity 24.8%; Pred. No. 1.8e-06;

Matches 55; Conservative 37; Mismatches 98; Indels 32; Gaps 8;

```
QY 210 CDKHTTCPELLGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKFNWY----- 263
   |  :  |||  |||  :  |||  :  |||  :  |||  :  |||  :  |||  :  |||  :  |||
Db 14 CTVSEASTYELTQPPSVSP-----GCTARITC---SGDALPKKAYWYIQKSGQ 60
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 264 VDGVEVHNAKTKPR--EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKA-----LPA 315
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 61 TPVLVIYDTERPSGIPERFSGSSSGTVALTLTSGAQVEDADYCYSSDSSGNHWVFG 120
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 316 PIEKTIKAKGQPR-EPQVYTLPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNGQP-E 373
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 121 GTKLTV---LGPKAAPSVTLFPSSSEELQANKATLVCLISDFYPGAVTVAMKADSPVK 177
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 374 NNYKTTPLVLDSDGSFLYSKLTVDKSRMGGGNVFSQSVME 415
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 178 AGVEITTPSKQSNKRYAASSTYLSLTPEQWKSHRSYSCVTHE 219
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
```

Search completed: August 3, 2004, 13:13:28
Job time : 38.1743 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:09 ; Search time 46.1899 Seconds
(without alignments)
2777.216 Million cell updates/sec

Title: SEQ4

Perfect score: 2414
Sequence: 1 MNRGVPRHLLVQLALLP.....DETCAQDGLDGLMTTDP 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 125 summaries

Database : A_Geneseq29Jan04:*

1: geneseqp19908:*\n2: geneseqp19908:*\n3: geneseqp20008:*\n4: geneseqp20018:*\n5: geneseqp20028:*\n6: geneseqp20038:*\n7: geneseqp20038:*\n8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2216	91.8	435	2	AAR26530 Sequence
2	2188.5	90.7	450	2	AAR37574 Human DID
3	2176	90.1	534	2	AAR26531 Sequence
4	2169	89.9	449	6	AAR37576 Human FDI
5	2126.5	88.1	631	3	AAP93009 Genetic C
6	2126.5	88.1	631	3	AAB19508 CD4-IgG1
7	2126.5	88.1	631	3	AAV51079 Human fus
8	2126.5	88.1	631	3	AAV59169 CD4-Ig fu
9	2122	87.9	432	2	AAR26782 CD4-gamma
10	2122	87.9	432	2	AAR46678 CD4-gamma
11	2122	87.9	432	2	AAH85079 Human CD4
12	2122	87.9	432	4	AAAB67332 CD4-gamma
13	2122	87.9	432	4	AAAB67332 Human CD4
14	2122	87.9	432	6	ABG71122 CD4-gamma
15	2085	86.4	530	2	AAR26783 CD4-IgG2
16	2085	86.4	530	3	AAH85080 CD4-IgG2
17	2085	86.4	530	4	AAAB67333 CD4-IgG2
18	2085	86.4	530	4	AAAB67333 Human CD4
19	2085	86.4	530	6	ABG71123 CD4-IgG2
20	2081	86.2	616	3	AAV51082 Human fus
21	2079.5	86.1	729	1	AAAP93008 Genetic C
22	2079.5	86.1	729	3	AAAB19507 CD4-IgG1
23	2079.5	86.1	729	3	AAV59168 CD4-Ig fu
24	2075	86.0	530	2	AAR46679 CD4-IgG2

26	2069.5	85.7	729	3	AAV51078 Human fus
27	2056	85.2	614	1	AAP93012 Genetic C
28	1999	82.8	616	3	AAAB19511 CD4-IgG1
29	1647.5	68.2	410	2	AAW35860 Human CD4
30	1546	64.0	343	2	AAAR20634 ChimERIC
31	1338.5	55.4	254	2	AAAR89441 IgG1 hing
32	1332.5	55.2	254	2	AAAR76667 IgG1 hing
33	1327.5	55.0	400	7	ADD13790 Plasmid p
34	1315	54.5	401	7	ADD13781 Plasmid p
35	1287.5	53.3	437	6	ABJ37104 Concatame
36	1287	53.3	461	2	AAAR42162 Anti-HIV
37	1284.5	53.2	449	5	AAO18400 Mature hu
38	1284.5	53.2	459	2	AAAR42066 Human ant
39	1282.5	53.1	446	2	AAW05829 Humanised
40	1282.5	53.1	617	6	ABJ37106 Concatame
41	1282.5	53.1	617	6	ABJ37108 Concatame
42	1282	53.1	582	4	AAAB81991 Gangliosid
43	1280.5	53.0	461	4	AAU07745 Human ant
44	1280.5	53.0	475	2	AAW11639 Human ant
45	1280	53.0	473	5	ABG70743 Mouse/hum
46	1280	53.0	476	2	AAW88464 Monoclonal
47	1279.5	53.0	448	5	AAW49203 Humanised
48	1279.5	53.0	452	2	AAV30201 Heavy cha
49	1279	53.0	470	3	AAV44721 Human imm
50	1279	53.0	492	7	ADD25783 Binding d
51	1278.5	53.0	581	4	AAAB81972 Gangliosid
52	1277	52.9	543	7	ADD25784 Binding d
53	1275.5	52.8	461	6	AAAR39847 Hu266 N56
54	1275.5	52.8	461	6	AAAR39843 Hu266 N56
55	1275.5	52.8	461	6	AAAR39843 Hu266 N56
56	1275.5	52.8	461	6	AAAR39843 Hu266 N56
57	1274.5	52.8	461	6	AAAR39843 Hu266 N56
58	1274.5	52.8	461	6	AAAR39843 Hu266 N56
59	1274.5	52.8	461	6	AAAR39843 Hu266 N56
60	1274.5	52.8	461	6	AAAR39843 Hu266 N56
61	1274	52.8	468	5	AAE27928 Human CSE
62	1274	52.8	468	5	AAE27928 Human CSE
63	1273.5	52.8	444	6	AAE35327 Antibody
64	1273.5	52.8	444	6	AAE35327 Antibody
65	1273.5	52.8	444	6	AAE35327 Antibody
66	1273.5	52.8	444	6	AAE35327 Antibody
67	1273.5	52.8	444	6	AAE35327 Antibody
68	1273.5	52.8	444	6	AAE35327 Antibody
69	1273.5	52.8	444	6	AAE35327 Antibody
70	1273	52.7	592	4	AAH83838 Amilo aci
71	1272.5	52.7	465	4	AADE64199 L2HCF pr
72	1272.5	52.7	465	4	AADE64199 L2HCF pr
73	1271.5	52.7	453	2	AAV50151 ChimERIC
74	1271.5	52.7	499	5	ABG31025 Antibody
75	1271.5	52.7	499	7	ADD25877 Binding d
76	1271.5	52.7	499	7	ADD25877 Binding d
77	1271	52.7	468	6	ABP58275 Humanised
78	1271	52.7	472	2	AAAR93166 Anti-rhes
79	1271	52.7	476	2	AAW01822 Primatise
80	1271	52.7	476	2	AAW01822 Primatise
81	1271	52.7	476	2	AAW01822 Primatise
82	1271	52.7	476	2	AAW01822 Primatise
83	1270.5	52.6	472	6	ABP58289 Humanised
84	1270.5	52.6	582	4	AAAB81987 Gangliosid
85	1270	52.6	619	5	AAU87089 Stiglec-BM
86	1269.5	52.6	744	3	AAV94408 Human VCA
87	1269.5	52.6	744	3	AAV94408 Human VCA
88	1268	52.5	634	6	ABH82300 CD19-zeta
89	1267.5	52.5	442	6	ABR33465 Humanised
90	1267.5	52.5	442	6	ABR33465 Humanised
91	1267.5	52.5	442	6	ABR33465 Humanised
92	1266	52.4	470	5	ABR81109 Anti-tlrs
93	1266	52.4	470	5	ABR81109 Anti-tlrs
94	1266	52.4	470	5	ABR81109 Anti-tlrs
95	1266	52.4	470	5	ABR81109 Anti-tlrs
96	1266	52.4	470	5	ABR81109 Anti-tlrs
97	1266	52.4	470	5	ABR81109 Anti-tlrs
98	1265.5	52.4	454	2	AAAR30774 H52H4-160

```

99 1265.5 52.4 468 2 AAW85689
100 1265.5 52.4 711 2 AAW85692
101 1265 52.4 448 6 ABR55871
102 1265 52.4 449 2 AAR43339
103 1265 52.4 449 2 AAW49816
104 1265 52.4 453 6 ABR58287
105 1265 52.4 462 4 AAY97592
106 1265 52.4 462 5 ABR52445
107 1265 52.4 466 2 AAR24812
108 1265 52.4 470 5 AAE27923
109 1265 52.4 470 6 ABR88232
110 1265 52.4 567 5 AAY97597
111 1265 52.4 567 5 ABR52442
112 1265 52.4 698 5 AAU81012
113 1265 52.4 698 6 AA016237
114 1264.5 52.4 445 2 AAW69316
115 1264.5 52.4 445 2 ABR58273
116 1264 52.4 451 5 AAB47726
117 1264 52.4 466 2 AAR40750
118 1264 52.4 567 4 AAY97593
119 1264 52.4 567 5 ABR52446
120 1263.5 52.3 450 4 AAE10517
121 1263.5 52.3 450 4 AAE10509
122 1263.5 52.3 450 5 ABR66608
123 1263.5 52.3 450 5 ABR66590
124 1263.5 52.3 450 5 ABR66610
125 1263.5 52.3 450 5 ABR66588

```

ALIGNMENTS

RESULT 1

AAR26530

ID AAR26530 standard; protein; 435 AA.

XX AAR26530;

DT 25-MAR-2003 (revised)

DT 28-JAN-1993 (first entry)

DE Sequence of one chain of a CD4-gamma 1 chimeric heavy chain homodimer.

XX CD4-gamma 1 chimeric heavy chain homodimer; expression vector; HIV;

KW therapy; diagnostic agent; inhibition.

XX OS Synthetic.

XX Key Location/Qualifiers

FT 1..204

FT /label= CD4

FT /note= "1..25 = preregion"

FT Region 205..219

FT /label= hinge

FT Region 220..329

FT /label= CH2

FT Region 330..436

FT /label= CH3

XX W09213559-A1.

XX 20-AUG-1992.

XX 10-FEB-1992; 92MO-US001152.

XX 08-FEB-1991; 91US-00654205.

XX (PROG-) PROGENICS PHARM INC.

XX Beauclry GA, Maddon PJ;

XX WPI; 1992-299758/36.

XX N-PSDB; AAO27830.

```

XX CD4-gamma 1 chimeric heavy chain homo-dimer and its expression vector -
PT for preventing and treating HIV infection useful as a diagnostic agent.
XX Example; Fig 3; 88pp; English.
XX Human CD4 cDNA was excised from pSP6T4 and cloned into M13mp18. The 2 kb
CC PetI/PstI fragment from pBR lambda1 contg. the human lambda 1 heavy
CC chain gene (contg. the hinge, CH2 and CH3 exons) was isolated and cloned
CC into the BAP-treated M13mp18/CD4 vector. To obtain a CD4-lambda 1
CC chimeric heavy chain gene, oligonucleotide-mediated site-directed
CC mutagenesis was performed to juxtapose the CD4 and lambda 1 heavy chain
CC DNA sequences, ligating the CD4 sequence in frame to the hinge exon. The
CC DNA was then cloned into pCDNA-1 to produce CD4-IgG1-pcDNA1 (ATCC 40951).
XX (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 435 AA:
SQ
Query Match 91.8%; Score 2216; DB 2; Length 435;
Best Local Similarity 97.7%; Pred. No. 3e-117;
Matches 425; Conservative 0; Mismatches 4; Indels 6; Gaps 2;
QY 2 NRGVPRHLLLVQLALPPATGANKVVGKGGDTVELTCTAQQKSIOFHWKNSNQIKI 61
Db 1 NRGVPRHLLLVQLALPPATGKRVVVGKGGDTVELTCTAQQKSIOFHWKNSNQIKI 60
QY 62 LGNGSFLTKGPSKLNDRADRSRSLMDQGNFLLIKVLTIEDSDTYICEVEDQKEVQL 121
Db 61 LGNGSFLTKGPSKLNDRADRSRSLMDQGNFLLIKVLTIEDSDTYICEVEDQKEVQL 120
QY 122 VFGILTANSDTHLLQGQSLTTLTLESPGSPSVQCRSPRGKNIQGGKTLVSQLELDQSGT 181
Db 121 VFGILTANSDTHLLQGQSLTTLTLESPGSPSVQCRSPRGKNIQGGKTLVSQLELDQSGT 180
QY 182 WTCTVLONQKKVEFKIDIVCPAPEPKSCDKITTC-----PELLGGSVTLFPPKPDTL 236
Db 181 WTCTVLONQKKVEFKIDIVL-APEPKSCDKITTCPECPAPELGGSVFLFPPKPDTL 239
QY 237 MISRTPEVTCVVDVSHEDPEVFNMYVDGVEVHNATKPREBOYNSTYVWSVLTVLHQ 296
Db 240 MISRTPEVTCVVDVSHEDPEVFNMYVDGVEVHNATKPREBOYNSTYVWSVLTVLHQ 299
QY 297 DWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKG 356
Db 300 DWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKG 359
QY 357 FYPSDIAVWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRMQQGVFSSGVMHEA 416
Db 360 FYPSDIAVWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRMQQGVFSSGVMHEA 419
QY 417 LHNHYTQKSLSLSPG 431
Db 420 LHNHYTQKSLSLSPG 434

```

RESULT 2

AAE37574

ID AAE37574 standard; protein; 450 AA.

XX AAE37574;

XX 27-AUG-2003 (first entry)

XX Human D1D2-Ig alphatp fusion protein.

XX Human, CD4; cluster of differentiation factor 4; immunoglobulin; Ig;

XX human immunodeficiency virus; gene therapy; vaccine; HIV-1 infection; D1;

XX D2; alpha tailpiece; alphatp; fusion protein.

XX Homo sapiens.

XX W02003040311-A2.

PD 15-MAY-2003.
 XX 24-OCT-2002; 2002WO-US034393.
 PF 25-OCT-2001; 2001US-0346231P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Arthos J, Cicala C, Fauci AS;
 PI WPI; 2003-441545/41.
 XX DR N-PSDB; AAD29113.
 XX
 PT New CD4 polypeptide ligated at its C-terminus with a portion of an
 PT immunoglobulin, useful for preparing a composition for treating or
 PT preventing HIV-1 infection.
 XX
 PS Example 1; Page 47; 100pp; English.
 XX
 CC The invention relates to a CD4 (cluster of differentiation factor 4)
 CC polypeptide ligated at its C-terminus with a portion of an immunoglobulin
 CC (Ig) comprising a hinge region and a constant domain of a mammalian Ig
 CC heavy chain. The polypeptide comprises a tailpiece from the C-terminus of
 CC the heavy chain of an IgA or IgM antibody. Polypeptides of the invention
 CC are useful for preparing a composition for treating or preventing human
 CC immunodeficiency virus (HIV)-1 infection. The invention is useful in gene
 CC therapy and also in the preparation of vaccines. The present sequence is
 CC a fusion protein which comprises a human IgA alpha tailpiece (alphanap), a
 CC human IgG constant region comprising a hinge, a CH2 and CH3 region and a
 CC human CD4 DID2 domain
 XX
 SQ Sequence 450 AA;

Query Match 90.7%; Score 2188.5; DB 6; Length 450;
 Best Local Similarity 93.2%; Pred. No. 1.1e-115;
 Matches 423; Conservative 2; Mismatches 8; Indels 21; Gaps 3;

QY 1 MNRGVFPHLLVLTQALPPAATGKNVYLGKKDPTVELTCTASQKSIQPHMKNSNOIK 60
 DB 1 MNRGVFPHLLVLTQALPPAATGKNVYLGKKDPTVELTCTASQKSIQPHMKNSNOIK 60
 QY 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEYQL 120
 DB 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEYQL 120
 QY 121 LVFGILTANSDTHLLQGQSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGILTANSDTHLLQGQSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWTCVVLQOKKVEFEFIDIVPCPAPRPKSCDKTHNC-----PELLGSPVFLFPPRPKDT 235
 DB 181 TWTCVVLQOKKVEFEFIDIVL-----ASADTHHTCTCPCPABELLGGPVEFLFPPRPKDT 235
 QY 236 LMISRTPEVTCVVVDVSHEDPEVKFMVYDGVENNAKTRPEEQNSTYRVVSVLTVLH 295
 DB 236 LMISRTPEVTCVVVDVSHEDPEVKFMVYDGVENNAKTRPEEQNSTYRVVSVLTVLH 295
 QY 236 QDWLNGKEYKCKVSNKALPAPIEKTTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK 355
 DB 236 QDWLNGKEYKCKVSNKALPAPIEKTTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK 355
 QY 356 GYPSDIAVEMESNQPENNYKTTPPVLDSDGSFLYSKLTVDKSRMKGNSFYCSVWME 415
 DB 356 GYPSDIAVEMESNQPENNYKTTPPVLDSDGSFLYSKLTVDKSRMKGNSFYCSVWME 415
 QY 416 ALAHNYTQKSLSPG-----IQLDETC 438
 DB 416 ALAHNYTQKSLSPG-----IQLDETC 438
 QY 416 ALAHNYTQKSLSPG-----IQLDETC 438
 DB 416 ALAHNYTQKSLSPG-----IQLDETC 438

RESULT 3
 AAR26531
 ID AAR26531 standard; protein; 534 AA.

XX AAR26531;
 AC 25-MAR-2003 (revised)
 XX 28-JAN-1993 (first entry)
 DT Sequence of CD4-IgG1 chimeric heavy chain heterotetramer.
 XX
 DE CD4-gamma 1 chimeric heavy chain homodimer; expression vector; HIV;
 XX therapy; diagnostic agent; inhibition.
 XX Synthetic.
 XX
 OS Key Location/Qualifiers
 XX FH Region 205..302
 XX FT Region /label= CH1
 XX FT Region 303..317
 XX FT Region /label= hinge
 XX FT Region 318..427
 XX FT Region /label= CH2
 XX FT Region 428..534
 XX FT Region /label= CH3
 XX
 XX WO9213559-A1.
 XX 20-AUG-1992.
 XX 10-FEB-1992; 92WO-US001152.
 XX 08-FEB-1991; 91US-00654205.
 XX (PROG-) PROGENICS PHARM INC.
 XX Beaudry GA, Maddon PJ;
 FI WPI; 1992-299758/36.
 DR N-PSDB; AAQ27631.
 XX
 PT CD4-gamma 1 chimeric heavy chain homo-dimer and its expression vector -
 PT for preventing and treating HIV infection useful as a diagnostic agent.
 XX
 PS Example; Fig 4; 88pp; English.
 XX
 CC The human CD4 cDNA is excised from the plasmid pSP6T4 and cloned into
 CC M13mp18. In order to excise a fragment containing the CH1 exon of the
 CC human gamma 1 heavy chain gene, the plasmid pBR gamma 1 is digested with
 CC SacII, and the SacII sites are then made flush using T4 DNA polymerase.
 CC The fragment containing the CH1 exon is then purified and ligated to the
 CC M13mp18 (CD4) vector. Oligonucleotide-mediated site-directed mutagenesis
 CC is then performed to juxtapose the CD4 and CH1 sequences in frame. The
 CC CD4-CH1 chimeric gene is then linearized and ligated to the pScl-Pscl DNA
 CC fragment of the plasmid pBR gamma 1 containing the hinge, CH2, and CH3
 CC exons of the human gamma 1 chain gene designated CD4-IgG1HC-pRCMV (ATCC
 CC 75192). (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 534 AA;

Query Match 90.1%; Score 2176; DB 2; Length 534;
 Best Local Similarity 80.3%; Pred. No. 6.7e-115;
 Matches 428; Conservative 0; Mismatches 3; Indels 102; Gaps 4;

QY 1 MNRGVFPHLLVLTQALPPAATGKNVYLGKKDPTVELTCTASQKSIQPHMKNSNOIK 60
 DB 1 MNRGVFPHLLVLTQALPPAATGKNVYLGKKDPTVELTCTASQKSIQPHMKNSNOIK 60
 QY 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEYQL 120
 DB 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEYQL 120
 QY 121 LVFGILTANSDTHLLQGQSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGILTANSDTHLLQGQSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 121 LVFGILTANSDTHLLQGQSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGILTANSDTHLLQGQSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180

```

QY 181 TWCTVLONOKKVEFKIDIV-----PC-----PAP 205
DB 181 TWCTVLONOKKVEFKIDIVLAFSTKQSVFPLAPSPKSTSGGTALGCLVXDYPPEP 240
QY 206 ----- 205
DB 241 VTSWNSGALTSVHTFPFAVLQSSGLYSLSVVTVPSSSLGTQTYICNVNHPKNTKVDK 300
QY 206 --EPKSCDKHTYC-----PELLGSPSVFLFPKPKDPTMISRTPEVTCVWVDVSHEDDEV 258
DB 301 KVEPSSCDKHTCTCPCPAPBELLGSPSVFLFPKPKDPTMISRTPEVTCVWVDVSHEDDEV 360
QY 259 KENVYVGVENVNAKTKPREEQYNSTYRVSVLTVLHODWLNKCKKCVSNKALPAPIE 318
DB 361 KENVYVGVENVNAKTKPREEQYNSTYRVSVLTVLHODWLNKCKKCVSNKALPAPIE 420
QY 319 KTISAKQOPRPQYVTLTPSRDELTKQVSLTCLVKGFPYPSDIAVEMESNGQPPNNYKT 378
DB 421 KTISAKQOPRPQYVTLTPSRDELTKQVSLTCLVKGFPYPSDIAVEMESNGQPPNNYKT 480
QY 379 TPPLVDSGSPFLYSKLTVDKSRMOQGNVFCSVNHEALHNHYTKSLSPG 431
DB 481 TPPLVDSGSPFLYSKLTVDKSRMOQGNVFCSVNHEALHNHYTKSLSPG 533

RESULT 4
AAE37576
ID AAE37576 standard; protein; 449 AA.
AC AAE37576;
DT 27-AUG-2003 (first entry)
DE Human FD1D2-Ig alphap2 fusion protein variant.
XX
XX Human, CD4; cluster of differentiation factor 4; immunoglobulin, Ig;
XX human immunodeficiency virus; gene therapy; vaccine; HIV-1 infection; D1;
XX D2; alpha tailpiece; alphap2; fusion protein; muten; variant; mutant.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 218 /note= "Wild type Glu substituted with Pro"
XX FT Misc-difference 219 /note= "Wild type Leu substituted with Val"
XX FT Misc-difference 221 /note= "Wild type Gly substituted with Ala"
XX
XX WO2003040311-A2.
XX
XX 15-MAY-2003.
XX
XX 24-OCT-2002; 2002WO-US034393.
XX
XX 25-OCT-2001; 2001US-0346231P.
XX
XX (US98 ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Archos J, Cicala C, Fauci AS;
XX
XX MPI; 2003-441545/41.
XX DR N-PSDB; ACC82877.
XX
XX New CD4 polypeptide ligated at its C-terminus with a portion of an
XX immunoglobulin, useful for preparing a composition for treating or
XX preventing HIV-1 infection.
XX
XX Example 11; Page 67; 100pp; English.
XX
XX The invention relates to a CD4 (cluster of differentiation factor 4)
XX polypeptide ligated at its C-terminus with a portion of an immunoglobulin
XX (Ig) comprising a hinge region and a constant domain of a mammalian Ig

```

```

CC heavy chain. The polypeptide comprises a tailpiece from the C-terminus of
CC the heavy chain of an IgA or IgM antibody. Polypeptides of the invention
CC are useful for preparing a composition for treating or preventing human
CC immunodeficiency virus (HIV)-1 infection. The invention is useful in gene
CC therapy and also in the preparation of vaccines. The present sequence is
CC a fusion protein variant (G218P/L219V/220delA/G221A) which comprises a
CC human IgA alpha tailpiece (alphap2), a human IgG2 constant region
CC comprising a hinge, a CH2 and CH3 region and a human CD4 D1D2 domain.
CC This variant protein is also referred to as mutant F.
CC
XX Sequence 449 AA:
SQ
Query Match 89.9%; Score 2169; DB 6; Length 449;
Beat Local Similarity 92.5%; Pred. No. 1,4e-114;
Matches 419; Conservative 3; Mismatches 11; Indels 20; Gaps 3;

QY 1 MNRGVPRHLLVLQLLPLPAATQGNVYLGKKDPTVELCTASQKSIQFMHNSNQIK 60
DB 1 MNRGVPRHLLVLQLLPLPAATQGNVYLGKKDPTVELCTASQKSIQFMHNSNQIK 60
QY 61 ILNGQSFLLTKGSPKLNDRADSRSLMDQGNFLLIKNLKI ESDTYICEVEDQKEEVQL 120
DB 61 ILNGQSFLLTKGSPKLNDRADSRSLMDQGNFLLIKNLKI ESDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTVLONOKKVEFKIDIVL-----ASADKHTHTCPCPAPVAGSVFLFPKPKDPTL 235
DB 181 TWCTVLONOKKVEFKIDIVL-----ASADKHTHTCPCPAPVAGSVFLFPKPKDPTL 235
QY 237 MISRTPEVTCVWVDVSHEDDEVKENVYVGVENVNAKTKPREEQYNSTYRVSVLTVLHQ 296
DB 237 MISRTPEVTCVWVDVSHEDDEVKENVYVGVENVNAKTKPREEQYNSTYRVSVLTVLHQ 296
QY 236 MISRTPEVTCVWVDVSHEDDEVKENVYVGVENVNAKTKPREEQYNSTYRVSVLTVLHQ 295
DB 236 MISRTPEVTCVWVDVSHEDDEVKENVYVGVENVNAKTKPREEQYNSTYRVSVLTVLHQ 295
QY 297 DWLNKGEYKCKVSNKALPAPIEKTISSAKQOPRPQYVTLTPSRDELTKQVSLTCLVKG 356
DB 297 DWLNKGEYKCKVSNKALPAPIEKTISSAKQOPRPQYVTLTPSRDELTKQVSLTCLVKG 355
QY 357 FYPSDIAVEMESNGQPPNNYKTTPPLVDSGSPFLYSKLTVDKSRMOQGNVFCSVNHEA 416
DB 357 FYPSDIAVEMESNGQPPNNYKTTPPLVDSGSPFLYSKLTVDKSRMOQGNVFCSVNHEA 415
QY 417 LHNHYTKSLSPG-----LQLD ETC 438
DB 416 LHNHYTKSLSPG-----LQLD ETC 448

RESULT 5
AAP93009
ID AAP93009 standard; protein; 631 AA.
AC AAP93009;
DT 25-MAR-2003 (revised)
DT 02-NOV-1992 (first entry)
DE Genetic construct which encodes CD4 linked to human IgG1 at the Bsp site
DE upstream of the hinge region (fusion protein CD4B-gamma-1).
XX
XX Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
XX diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
XX Homo sapiens.
XX
XX EP325262-A.
XX
XX 26-JUL-1989.
XX
XX 20-JAN-1989; 89EP-00100913.
XX
XX 22-JAN-1988; 88US-00147351.

```

```

XX (GEO) GEN HOSPITAL CORP.
XX PA
XX Seed B:
XX WPI, 1989-214472/30.
XX DR N-PSDB; AAN90357.
XX
XX Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
XX PT infections or detecting HIV or SIV in sample.
XX
XX Example: Table 2, Page 24-33; 68pp; English.
XX
XX The fusion protein genes of the invention pref. comprise cDNA sequences
XX which encode CD4 or a fragment which binds gp120 ligated to an expression
XX plasmid which encodes an antibody in which the variable region of the
XX gene has been deleted (see WO87-02671). The CD4 portion of the fusion
XX protein may comprise the complete CD4 sequence, the 370 AA extracellular
XX region and the membrane spanning domain, or the extracellular region. The
XX Ig heavy chain is pref. from IGM, IgG1 or IgG3. The following are
XX specifically claimed: fusion proteins CD4H-gamma-1, CD4mu, CD4pmu, CD4E-
XX gamma1, and CD4mmu (No. 67608), pCD4F-gamma (No. 67609) and pCD4E-gamma-1
XX (No. 67610). The plasmid containing (pCD4E-gamma-1) has been deposited in
XX E. coli (MC1061/P3) at the ATCC under accession number 67610. (Updated on
XX 25-MAR-2003 to correct PA field.)
XX
XX Sequence 631 AA:
SQ
Query Match      88.1%; Score 2126.5; DB 1; Length 631;
Best Local Similarity 67.8%; Pred. No. 4.9e-112;
Matches 427; Conservative 0; Mismatches 4; Indels 199; Gaps 3;

QY 1 NMRGVFRRLLLVLTQALLPAATQGNKVVLGKGGDVELTCTASQKSIQFHMKNQIK 60
DB 1 NMRGVFRRLLLVLTQALLPAATQGNKVVLGKGGDVELTCTASQKSIQFHMKNQIK 60
QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLITKNKIKDSPTVYICEVDOKEEYOL 120
DB 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLITKNKIKDSPTVYICEVDOKEEYOL 120
QY 121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180
DB 121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180
QY 181 TWTCTVLQOKKVEFKIDIV-----PCPAP----- 200
DB 181 TWTCTVLQOKKVEFKIDIVLAFOKASSIVYKKEGQVEFSFPLAFTVEKLTGSGELMW 240
QY 201 -----PCPAP----- 205
DB 241 QAEKASSKSMITPDLKKEVSKVYQDPKIQMGKKLPLHLTLFQALPQVAGSGLTLTA 300
QY 206 ----- 205
DB 301 LEAKGKLHQEVNLYVMRATQLOKVLTCFVWGPTSPKMLSLKLENKAKVSKREKPVWV 360
QY 206 -----EPKSCDKTHTC-----PELIG 221
DB 361 INPEAGMQCLSDSGQVLLBSNIKVLPYWSPTPVHADPEEPKSCDKTHTCPCPAPPELIG 420
QY 222 GPSVFLFPPKPKDITLMISRTPEVTCVVDVSHEDPEVKENYVDGVEYNATATKREBOY 281
DB 421 GPSVFLFPPKPKDITLMISRTPEVTCVVDVSHEDPEVKENYVDGVEYNATATKREBOY 480
QY 282 NSTYRVSVLTVLHODMLNGKEYKCVSNKALPAPIEKTISKAKQPREPOVYTLPPSRD 341
DB 481 NSTYRVSVLTVLHODMLNGKEYKCVSNKALPAPIEKTISKAKQPREPOVYTLPPSRD 540
QY 342 ELTKQVSLTCLVKGFYPSDIAVWESNQPENNYKTPPVYDSGSPFLYSKLTVDXSR 401
DB 541 ELTKQVSLTCLVKGFYPSDIAVWESNQPENNYKTPPVYDSGSPFLYSKLTVDXSR 600
QY 402 WQGNVFCGSVWHEALHNHYTQKSLSLSPG 431

```

```

DB 601 WQGNVFCGSVWHEALHNHYTQKSLSLSPG 630
|||||
RESULT 6
AAB19508
ID AAB19508 standard; protein; 631 AA.
XX
XX AAB19508;
AC
XX 09-JUN-2001 (first entry)
DT
XX
XX CD4-IgG1 fusion protein CH4Egamma1.
DE
XX
XX CD4: IgG1; human; CD4Egamma1; fusion protein; immunoglobulin; HIV; SIV;
XX gp120; therapy; diagnosis.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Protein 1..395
FT /note="CD4 extracellular region"
FT Protein 400..631
FT /note="IgG1 heavy chain"
XX
XX US6117656-A.
XX
XX 12-SEP-2000.
XX
XX 07-JUN-1995; 95US-00479353.
XX
XX 22-JAN-1988; 88US-00147351.
XX 23-JAN-1988; 89US-00289586.
XX 09-JUN-1992; 92US-00896781.
XX 12-APR-1993; 93US-00057952.
XX 04-FEB-1994; 94US-00191708.
XX
XX (GEO) GEN HOSPITAL CORP.
XX
XX Seed B:
XX
XX WPI: 2000-586558/55.
XX DR N-PSDB; AAS50661.
XX
XX CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
XX SIV.
XX
XX Example 1; Col 29-42; 39pp; English.
XX
XX The present sequence is that of fusion protein CD4Egamma1 comprising the
XX extracellular portion of CD4, which binds to HIV gp120, linked at its C-
XX terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
XX encoding CD4 was linked to IgG1 DNA at the Esp site upstream of the hinge
XX region (see AAS50661). Fusion protein CD4Egamma1 and a nucleic acid
XX encoding it are claimed. Also claimed are a vector comprising the nucleic
XX acid, and a method of producing the fusion protein in secreted form using
XX a transformed host cell. The fusion protein may further comprise a
XX therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
XX can be administered to an animal (including humans) for treatment of HIV
XX or SIV infection, and can also be used in assays for HIV or SIV, imaging
XX and tissue stains. IgG1 fusion proteins such as CD4Egamma1 provide both
XX complement-mediated and cell-mediated immunity
XX
XX Sequence 631 AA:
SQ
Query Match      88.1%; Score 2126.5; DB 3; Length 631;
Best Local Similarity 67.8%; Pred. No. 4.9e-112;
Matches 427; Conservative 0; Mismatches 4; Indels 199; Gaps 3;

QY 1 NMRGVFRRLLLVLTQALLPAATQGNKVVLGKGGDVELTCTASQKSIQFHMKNQIK 60
DB 1 NMRGVFRRLLLVLTQALLPAATQGNKVVLGKGGDVELTCTASQKSIQFHMKNQIK 60

```

```
QY 61 ILNGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDPHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDPHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCTVLONQKKVEFKIDIV----- 200
DB 181 TWTCTVLONQKKVEFKIDIV----- 200
QY 201 TWTCTVLONQKKVEFKIDIVLAPOKASSIYKKKEGQVRSFPLATVEKLTGSGELMW 240
DB 201 TWTCTVLONQKKVEFKIDIVLAPOKASSIYKKKEGQVRSFPLATVEKLTGSGELMW 240
QY 241 QABRASSKSNITFDLKNKEVSVKRVTPDKLQMGKLPYLHLTPQALPQYAGSGLTLTA 300
DB 241 QABRASSKSNITFDLKNKEVSVKRVTPDKLQMGKLPYLHLTPQALPQYAGSGLTLTA 300
QY 206 -----PCPAP----- 205
DB 206 -----PCPAP----- 205
QY 301 LEAKTGKLGQEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
DB 301 LEAKTGKLGQEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
QY 206 -----EPKSCDKTHTC-----PELLG 221
DB 206 -----EPKSCDKTHTC-----PELLG 221
QY 361 LNPEAGMWCCLSDSGVLLSNTIKVLPWTSTPVHADPEBPKSCDKTHTCPCPAPPELLG 420
DB 361 LNPEAGMWCCLSDSGVLLSNTIKVLPWTSTPVHADPEBPKSCDKTHTCPCPAPPELLG 420
QY 222 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 281
DB 222 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 281
QY 421 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 480
DB 421 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 480
QY 282 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 341
DB 282 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 341
QY 481 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 540
DB 481 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 540
QY 342 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 401
DB 342 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 401
QY 541 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 600
DB 541 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 600
QY 402 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 431
DB 402 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 431
QY 601 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 630
DB 601 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 630

RESULT 7
ID AAY51079 standard; protein; 631 AA.
XX AC AAY51079;
XX DT 23-MAR-2000 (first entry).
XX DE Human fusion protein CD4Egammal.
XX KW Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
XX KW anti-human immunodeficiency virus; CD4Egammal.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN US6004781-A.
XX PD 21-DEC-1999.
XX PE 04-FEB-1994; 94US-00191708.
XX PR 22-JAN-1988; 88US-00147351.
XX PR 23-JAN-1989; 89US-00299596.
XX PR 09-JUN-1992; 92US-00886781.
XX PR 12-APR-1993; 93US-00057952.
XX PA (GEMO) GEN HOSPITAL CORP.
XX PI Seed B;
XX DR WPI; 2000-085792/07.
```

```
DR N-PSDB; AAZ44062.
XX Fusion protein useful for the treatment of human immunodeficiency virus.
XX PS Example 1; Col 29-42; 39pp; English.
XX CC This invention describes a novel nucleic acid (I) encoding a fusion
XX CC protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
XX CC and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
XX CC chain (III). The products of the invention have anti-human
XX CC immunodeficiency virus (HIV) activity and are capable of binding to
XX CC gp120. The fusion protein is useful for treating human immunodeficiency
XX CC virus (HIV) or simian immunodeficiency virus (SIV). This sequence
XX CC represents the fusion protein CD4Egammal which is constructed from CD4
XX CC linked to human IgG1 upstream of the hinge region
XX SO Sequence 631 AA.

Query Match 88.1%; Score 2126.5; DB 3; Length 631;
Best Local Similarity 67.8%; Pred. No. 4,9e-112;
Matches 427; Conservative 0; Mismatches 4; Indels 199; Gaps 3;

QY 1 MNRGVPFRHLLVQLALPAAATQGNKVVLGKGDVVELCTASQKSIQFHMNSNOIK 60
DB 1 MNRGVPFRHLLVQLALPAAATQGNKVVLGKGDVVELCTASQKSIQFHMNSNOIK 60
QY 61 ILNGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDPHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDPHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCTVLONQKKVEFKIDIV----- 200
DB 181 TWTCTVLONQKKVEFKIDIV----- 200
QY 201 TWTCTVLONQKKVEFKIDIVLAPOKASSIYKKKEGQVRSFPLATVEKLTGSGELMW 240
DB 201 TWTCTVLONQKKVEFKIDIVLAPOKASSIYKKKEGQVRSFPLATVEKLTGSGELMW 240
QY 241 QABRASSKSNITFDLKNKEVSVKRVTPDKLQMGKLPYLHLTPQALPQYAGSGLTLTA 300
DB 241 QABRASSKSNITFDLKNKEVSVKRVTPDKLQMGKLPYLHLTPQALPQYAGSGLTLTA 300
QY 206 -----PCPAP----- 205
DB 206 -----PCPAP----- 205
QY 301 LEAKTGKLGQEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
DB 301 LEAKTGKLGQEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
QY 206 -----EPKSCDKTHTC-----PELLG 221
DB 206 -----EPKSCDKTHTC-----PELLG 221
QY 361 LNPEAGMWCCLSDSGVLLSNTIKVLPWTSTPVHADPEBPKSCDKTHTCPCPAPPELLG 420
DB 361 LNPEAGMWCCLSDSGVLLSNTIKVLPWTSTPVHADPEBPKSCDKTHTCPCPAPPELLG 420
QY 222 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 281
DB 222 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 281
QY 421 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 480
DB 421 GPSVFLPPPKPDITLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVHNAAKTKPREEQY 480
QY 282 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 341
DB 282 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 341
QY 481 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 540
DB 481 NSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPQVYTLPPSRD 540
QY 342 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 401
DB 342 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 401
QY 541 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 600
DB 541 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 600
QY 402 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 431
DB 402 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 431
QY 601 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 630
DB 601 WQGNVFSQSVMEHALHNHYTQKSLSLSPG 630

RESULT 8
ID AAY59169 standard; protein; 631 AA.
XX AC AAY59169;
XX DT 23-MAR-2000 (first entry).
```

```

DT 14-MAR-2000 (first entry)
XX
DE CD4-Ig fusion protein CD4Egammal.
XX
KM HIV, extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
XX secreted protein; SIV infection; medicament.
XX
OS Synthetic.
XX Homo sapiens.
XX
PN CA1340741-C.
XX
PD 14-SEP-1999.
XX
PF 20-JAN-1989; 89CA-00588749.
XX
PR 20-JAN-1989; 89CA-00588749.
XX
PA (GEHO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-063015/06.
XX
DR N-PSDB; AA248202.
XX
PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT the treatment of HIV or simian immunodeficiency virus infections.
XX
XX Example 1; Page 37-46; 89pp; English.
XX
CC The invention provides a fusion gene encoding a fusion protein that
CC comprises an extracellular CD4 DNA sequence or its fragment which binds
CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC the variable region has been replaced with the DNA sequence which encodes
CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC capable of being secreted. The fusion proteins are useful for treating
CC HIV or SIV infections in animals, preferably humans. They are also useful
CC for producing medicaments which can be used for treating HIV or SIV
CC infections in humans. The present sequence represents the fusion protein
CC CD4gamma1 where the CD4 is linked to human IgG1 at the Bsp site upstream
CC of the hinge region
XX
SQ Sequence 631 AA;
XX
Query Match 88.1%; Score 2126.5; DB 3; Length 631;
Best Local Similarity 67.8%; Pred. No. 4.9e-112;
Matches 427; Conservative 0; Mismatches 4; Indels 199; Gaps 3;

```

```

QY 206 -----EPKSCDKTHTC-----PEILG 221
DB 361 LNPEAGMWMQCLLSDSGCVLLESNIKVLPWTGSPVHADPEEPKSCDKTHTCPCAPPELLG 420
QY 222 GPSVFLPPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQY 281
DB 421 GPSVFLPPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPREEQY 480
QY 282 NSTYRVSVLTVLTHQDMINGKPKCKVSNKALPAPIEKTISKAKGQPREPPQVYTLPPSRD 341
DB 481 NSTYRVSVLTVLTHQDMINGKPKCKVSNKALPAPIEKTISKAKGQPREPPQVYTLPPSRD 540
QY 342 ELTKQVSLTCLVKGFPYSDIAVEWESNGQPENNYKTTIPVLDSDGSFFLYSKLTVDKSR 401
DB 541 ELTKQVSLTCLVKGFPYSDIAVEWESNGQPENNYKTTIPVLDSDGSFFLYSKLTVDKSR 600
QY 402 WQGNVVFSCSVNHEALHNHYTQKSLSLSPG 431
DB 601 WQGNVVFSCSVNHEALHNHYTQKSLSLSPG 630

RESULT 9
AAR26782
ID AAR26782 standard; protein; 432 AA.
XX
AC AAR26782;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 06-FEB-1993 (first entry)
XX
DE CD4-gamma2 chimeric heavy chain homodimer.
XX
KM homodimer; soluble CD4; T cell receptor; CD4 antigen; high recovery;
KM chimeric; increased serum half life; HIV infection; AIDS; ss.
XX
OS Homo sapiens.
OS Chimeric.
FH Key Location/Qualifiers
FT Domain 1..216
FT Domain /label= CD4
FT Domain 217..325
FT Domain /label= CH2
FT Domain 326..433
FT Domain /label= CH3
XX
PN WO9213947-A1.
XX
PD 20-AUG-1992.
XX
PF 10-FEB-1992; 92WO-US001143.
XX
PR 08-FEB-1991; 91US-00653684.
XX
PA (PROG-) PROGENICS PHARM INC.
XX
PI Beaudry GA, Maddon PJ;
XX
DR WPI; 1992-300034/36.
XX
DR N-PSDB; AAQ28088.
XX
PT CD4-gamma-2 and CD4-IgG2 chimera(s) and expression vectors - for
PT treatment, prevention and diagnosis of HIV infection.
XX
PS Claim 2; Fig 3; 90pp; English.
XX
CC This sequence represents a CD4-gamma2 chimeric heavy chain homodimer. It
CC was produced by expression of the coding mutagenised cDNA (produced as
CC described in AAQ28088) in Dhfr-CHO cells. The protein is efficiently
CC assembled intracellularly and effectively secreted from mammalian cells
CC pref. CHO, COS, or myeloma cells as a homodimer, enabling high recovery
CC and purification from the medium of cells expressing it. It possesses

```

CC increased serum half-life and has increased avidity for HIV cf. heavy
CC chain dimers. It can inhibit HIV infection of CD4+ cells and block the
CC spread of HIV infection within a patient. Attachment to a detectable
CC marker makes it useful in an assay for HIV or SIV infection, and it can
CC also be linked to toxins, eg diphtheria, Pseudomonas exotoxin A (domains
CC I or II) or the deglycosylated A-chain of ricin. (Updated on 25-MAR-2003
CC to correct PN field.) (Updated on 24-OCT-2003 to standardise OS field)

SQ Sequence 432 AA;

Query Match 87.9%; Score 2122; DB 2; Length 432;

Best Local Similarity 91.2%; Pred. No. 5,9e-112; Mismatches 2;

Matches 405; Conservative 8; Indels 26; Gaps 2;

QY 1 MNRGVPFRHLHLVQALALPPAATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPFRHLHLVQALALPPAATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
QY 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLOKQKVEFKIDIV-----PCPAPEPKSCDKHTCPELGGPSVFL 227
DB 181 TWTCVTLOKQKVEFKIDIVLAPEKCCVCECPPEPAP-----VAGPSVFL 227
QY 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVKFNMYDGVENVNAKTPREBOQNSTYRV 287
DB 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVKFNMYDGVENVNAKTPREBOQNSTYRV 287
QY 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVQFMNYYVGVENVNAKTPREBOQNSTYRV 287
DB 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVQFMNYYVGVENVNAKTPREBOQNSTYRV 287
QY 288 VSVLTCTVLOKQKVEFKIDIVLAPEKCCVCECPPEPAP-----VAGPSVFL 347
DB 288 VSVLTCTVLOKQKVEFKIDIVLAPEKCCVCECPPEPAP-----VAGPSVFL 347
QY 348 VSLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNV 407
DB 348 VSLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNV 407
QY 408 FSCSVMEALHNHYTQKSLSPG 431
DB 408 FSCSVMEALHNHYTQKSLSPG 431

RESULT 10

AAR46678 ID AAR46678 standard; protein; 432 AA.

XX AAR46678;

XX 25-MAR-2003 (revised)

DT 08-AUG-1994 (first entry)

XX CD4-gamma 2 chimeric heavy chain.

XX CD4; gamma; heavy chain; chimeric; chimeric; immunoconjugate; HIV;

KW human immunodeficiency virus; radionuclide; toxin; therapy; treatment;

XX imaging; detection; targeting.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 1..204

FT Region /label= CD4 Region.

FT Region /label= Hinge Region.

FT Region /label= CH2 Region.

FT Region /label= CH3 Region.

XX MO9403191-A1.

XX 17-FEB-1994.

XX 06-AUG-1993; 93MO-US007422.

XX 07-AUG-1992; 92US-00927931.

XX (PROG-) PROGENICS PHARM INC.

XX Allaway GP, Maddon PJ;

XX WPI; 1994-065392/08.

XX N-PSDB; AAG57750.

XX Non-peptidyl toxin or radionuclide and CD4-gamma 2 or CD4-IgG2

XX Immunoconjugates - used to kill HIV-infected cells and to image and

XX stage HIV infection.

XX Disclosure; Fig 3; 142pp; English.

CC A CD4-gamma 2 chimeric heavy chain homodimer is linked to a non-peptidyl
CC toxin or a gamma radiation-emitting radionuclide of low to moderate
CC cytotoxicity. The resulting immunoconjugate comprising the toxin can be
CC used to kill HIV infected cells and to treat HIV infected subjects to
CC reduce the population of HIV infected cells. It can also be used to
CC reduce the likelihood of infection. The immunoconjugate comprising the
CC radionuclide can be used to image HIV infected tissue, to calculate the
CC stage of HIV infection or the efficacy of an anti-HIV treatment using the
CC imaging technique and for determining the prognosis of an HIV infected
CC subject. (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 432 AA;

Query Match 87.9%; Score 2122; DB 2; Length 432;

Best Local Similarity 91.2%; Pred. No. 5,9e-112;

Matches 405; Conservative 8; Mismatches 5; Indels 26; Gaps 2;

QY 1 MNRGVPFRHLHLVQALALPPAATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPFRHLHLVQALALPPAATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
QY 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLOKQKVEFKIDIV-----PCPAPEPKSCDKHTCPELGGPSVFL 227
DB 181 TWTCVTLOKQKVEFKIDIVLAPEKCCVCECPPEPAP-----VAGPSVFL 227
QY 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVKFNMYDGVENVNAKTPREBOQNSTYRV 287
DB 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVKFNMYDGVENVNAKTPREBOQNSTYRV 287
QY 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVQFMNYYVGVENVNAKTPREBOQNSTYRV 287
DB 228 FPPKPKDTLMISRTPEVTCVAVDVSHEDPEVQFMNYYVGVENVNAKTPREBOQNSTYRV 287
QY 288 VSVLTCTVLOKQKVEFKIDIVLAPEKCCVCECPPEPAP-----VAGPSVFL 347
DB 288 VSVLTCTVLOKQKVEFKIDIVLAPEKCCVCECPPEPAP-----VAGPSVFL 347
QY 348 VSLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNV 407
DB 348 VSLTCLVKGFPYPSDIAVEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOQGNV 407
QY 408 FSCSVMEALHNHYTQKSLSPG 431
DB 408 FSCSVMEALHNHYTQKSLSPG 431

RESULT 11

AAV85079
ID AAV85079 standard; protein; 432 AA.
XX
AC AAV85079;
XX
DT 19-JUN-2000 (first entry)
XX
DE Human CD4-gamma 2 chimeric heavy chain homodimer amino acid sequence.
XX
KM CD4-gamma 2 chimeric heavy chain homodimer; immunodeficiency; treatment;
XX cytotoxic radionuclide; cell surface glycoprotein; prevent; infection;
XX cellular immune response interaction mediator; HIV interaction; staging;
XX prognosis; envelope glycoprotein burden; human.
OS Homo sapiens.
XX
XX US6034223-A.
XX
XX PD 07-MAR-2000.
XX
XX PF 07-JUN-1995; 95US-00477460.
XX
XX PR 07-AUG-1992; 92US-00927931.
XX 06-AUG-1993; 93WO-US007422.
XX PR 03-FEB-1995; 95US-00379516.
XX
XX (PROG-) PROGENICS PHARM INC.
XX
XX Allaway GP, Madden PJ;
XX
XX WPI; 2000-269502/23.
XX
XX DR N-PSDB; AA298655.
XX
XX PT New immunodeficiency virus infection, comprises radionuclide attached to
XX heterotetramer of CD4-immunoglobulin chimera.
XX
XX PS Disclosure; Fig 3; 58pp; English.
XX
XX This sequence represents the human CD4-gamma 2 chimeric heavy chain
XX immunodeficiency amino acid sequence. The invention relates to an
XX immunodeficiency comprising a cytotoxic radionuclide and a heterotetramer
XX of two heavy chains and two light chains. The cytotoxic radionuclide is
XX linked to either the heavy chains or the light chains, or to all four
XX chains, directly or through a bifunctional chelator. Both heavy chains
XX are chimeric CD4-Ig (immunoglobulin) G2 chains encoded by vector CD4-
XX IgG2HC-PRCCMV (ATCC 75193) and both light chains are chimeric CD4-Kappa
XX chains encoded by vector CD4-KLC-PRCCMV (ATCC 75194). CD4 is a non-
XX polymorphic cell surface glycoprotein that is expressed on the surface of
XX helper T lymphocytes, cells of the monocyte/macrophage lineage and
XX dendritic cells. CD4 associates with major histocompatibility complex
XX (MHC) class II molecules on the surface of antigen presenting cells to
XX mediate efficient cellular immune response interactions. In humans CD4 is
XX the target of interaction with the human immunodeficiency virus HIV. The
XX immunodeficiency is used to kill cells infected with HIV, and for treating
XX or preventing infection. It is also used for imaging HIV-infected tissues
XX (for staging or prognosis of infection, and for assessing efficacy of
XX treatments). The immunodeficiency is also used to determine the HIV
XX envelope glycoprotein burden, once determined, this information is used
XX in the staging and prognosis of HIV infected patients. The
XX immunodeficiency should be active against all strains of HIV (since the
XX CD4-gp120 interaction is essential for infection). The heterotetramers
XX are assembled intracellularly and secreted efficiently from mammalian
XX cells, allowing high recovery and purification from the culture medium.
XX They have longer half-life in serum and greater avidity than heavy chain
XX dimers
XX
XX Sequence 432 AA;

Query Match 87.9%; Score 2122; DB 3; Length 432;
Best Local Similarity 91.2%; Pred. No. 5.9e-112;
Matches 405; Conservative 8; Mismatches 5; Indels 26; Gaps 2;

QY 1 MNRGVPFRHLILLVQLALLPATOQNKVYLGKGGDTVELTCTAQQKSIOFHMKNSNOIK 60
DB 1 MNRGVPFRHLILLVQLALLPATOQNKVYLGKGGDTVELTCTAQQKSIOFHMKNSNOIK 60
QY 61 ILGNQGSFLTKGPPSLANDRADSRRLMDQGNFPLIINKLKIEDSTYICVEDQKEVQL 120
DB 61 ILGNQGSFLTKGPPSLANDRADSRRLMDQGNFPLIINKLKIEDSTYICVEDQKEVQL 120
QY 121 LVFGLTANSPTHLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGGKTLVSQLELODSG 180
DB 121 LVFGLTANSPTHLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGGKTLVSQLELODSG 180
QY 181 TWTCTVLONQKKVEFEKIDIV-----PCPAPEPKSCDKTHITCPBELLGPSVFL 227
DB 181 TWTCTVLONQKKVEFEKIDIVLAFAFKCCVCEPCPEPAP-----VAQPSVFL 227
QY 228 FPPKPKDTLMSIRTPETCVVVDVSHEDPEVKENNYVDGVEVNAKTKPREQVNSTRV 287
DB 228 FPPKPKDTLMSIRTPETCVVVDVSHEDPEVQFNMYVDGVEVNAKTKPREQVNSTRV 287
QY 288 VSVLTVLHODMNGEKYCKVSNKLLPAIEKTIKAGQPREPVYTLPSRDELTYNQ 347
DB 288 VSVLTVLHODMNGEKYCKVSNKLLPAIEKTIKAGQPREPVYTLPSRDELTYNQ 347
QY 348 VSLTCLVKGFPSPDIAYEWESNGCPENNYKTPPVLDSDGSFFLYSKLTVDSKRMQGNV 407
DB 348 VSLTCLVKGFPSPDIAYEWESNGCPENNYKTPPVLDSDGSFFLYSKLTVDSKRMQGNV 407
QY 408 FSCGYMHEALHNHYTQKSLSLSPG 431
DB 408 FSCGYMHEALHNHYTQKSLSLSPG 431

RESULT 12

AAV85079
ID AAV85079 standard; protein; 432 AA.

XX
XX AAV85079;
XX

XX DT 23-APR-2001 (first entry)
XX

XX DE CD4-gamma2 chimeric heavy chain homodimer protein.
XX

XX KM Immunodeficiency; chelator; chimeric; HIV; human immunodeficiency virus.
XX

XX OS Homo sapiens.
XX

XX PN US6177549-B1.
XX

XX PD 23-JAN-2001.
XX

XX PF 10-JUN-1999; 99US-00329916.
XX

XX PR 07-AUG-1992; 92US-00927931.
XX

XX PR 06-AUG-1993; 93WO-US007422.
XX

XX PR 03-FEB-1995; 95US-00379516.
XX

XX PR 07-JUN-1995; 95US-00477460.
XX

XX PA (PROG-) PROGENICS PHARM INC.
XX

XX PI Madden PJ, Allaway GP;
XX

XX DR WPI; 2001-158582/16.
XX

XX PT Immunodeficiency for treating human immunodeficiency virus-infected
XX subject, consists of cytotoxic radionuclide linked to heterotetramer
XX comprising two chimeric CD4-IgG2 heavy chains and two chimeric CD4-
XX kappa/lambda chains.
XX

XX PS Disclosure; Fig 3; 43pp; English.
XX

XX CC The present invention relates to an immunodeficiency, comprising a
XX cytotoxic radionuclide linked, directly or via a bifunctional chelator,

CC to a heterotetramer having two chimeric CD4-IgG2 heavy chains encoded by
 CC an expression vector CD4-IgG2HC-prcCMV and two chimeric CD4-kappa light
 CC chains encoded by an expression vector CD4-kLC-prcCMV. The invention is
 CC useful for killing human immunodeficiency virus (HIV)-infected cells, for
 CC the treatment and prevention of infection with HIV
 XX
 SQ Sequence 432 AA;

Query Match 87.9%; Score 2122; DB 4; Length 432;
 Best Local Similarity 91.2%; Pred. No. 5.9e-112;
 Matches 405; Conservative 8; Mismatches 5; Indels 26; Gaps 2;

QY 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
 DB 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
 QY 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
 DB 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
 QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180
 QY 181 TWCTVVLONOKKVEFKIDIV-----PCPAPEPKSCDKHTTCEPLLGGPSVFL 227
 DB 181 TWCTVVLONOKKVEFKIDIVLAPEKCCVCEPCPCPAP-----VAGPSVFL 227
 QY 228 FPPKPKDTLMSRPEVTCVVVDVSHEDPEVKFMNVYDGVENNAKTKPREBOYNSTRY 287
 DB 228 FPPKPKDTLMSRPEVTCVVVDVSHEDPEVKFMNVYDGVENNAKTKPREBOYNSTRY 287
 QY 288 VSVLTVLHQMNLNGKEYKCVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ 347
 DB 288 VSVLTVLHQMNLNGKEYKCVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ 347
 QY 348 VSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSGSPFLYSKLTVDKSRMOQGNV 407
 DB 348 VSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSGSPFLYSKLTVDKSRMOQGNV 407
 QY 408 FSCSVMEALHNHYTQKSLSLSPG 431
 DB 408 FSCSVMEALHNHYTQKSLSLSPG 431

RESULT 13

AAB80883
 ID AAB80883 standard; protein; 432 AA.

XX AAB80883;
 AC
 XX 29-MAY-2001 (first entry)
 DT
 XX Human CD4-gamma2 chimeric heavy chain homodimer.
 DE
 XX Human; Anti-HIV; CD4-IgG2 chimeric heterotetramer;
 KW Immunoglobulin gamma 2.
 XX
 OS Homo sapiens.
 XX
 XX US6187748-B1.
 PN
 XX 13-FEB-2001.
 PD
 XX 07-JUN-1995; 95US-00485372.
 PF
 XX 08-FEB-1991; 91US-00653684.
 PR 10-FEB-1992; 92MO-US001143.
 PR 08-DEC-1992; 92US-00960440.
 XX
 PA (PROG-) PROGENICS PHARM INC.
 XX
 FI Maddon PJ, Beaudry GA;

XX WPI; 2001-264981/27.
 DR N-PSDB; AAF77829.
 XX
 XX Inhibiting human immunodeficiency virus (HIV) infection of a CD4+ cell,
 PT or treating a subject having CD4+ cells infected with HIV involves using
 PT CD4-IgG2 chimeric heterotetramer to form a complex with the HIV.
 XX
 PS Disclosure; Fig 3; 55pp; English.

CC The present invention relates to a method for inhibiting infection of a
 CC CD4+ cell by HIV. The method comprises contacting the HIV with a CD4-IgG2
 CC chimeric heterotetramer to form a complex with the HIV (CD4 = cluster of
 CC glycoprotein that is expressed primarily on the surface of T cells. In
 CC man, CD4 is the target of interaction with HIV. The heterotetramer has
 CC two heavy and two light chains which are encoded by expression vectors
 CC CD4-IgG2HC-prcCMV (V1) and CD4-kLC-prcCMV (V2), respectively. The method
 CC is used to inhibit infection of a CD4+ cell by a HIV and to prevent CD4+
 CC cells of a subject from becoming infected with HIV. The method is also
 CC useful for treating a subject having CD4+ cells infected with HIV. The
 CC present sequence is human fusion protein: CD4-gamma2 chimeric heavy chain
 CC homodimer. This sequence was used in the method of the present invention
 XX
 SQ Sequence 432 AA;

Query Match 87.9%; Score 2122; DB 4; Length 432;
 Best Local Similarity 91.2%; Pred. No. 5.9e-112;
 Matches 405; Conservative 8; Mismatches 5; Indels 26; Gaps 2;

QY 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
 DB 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
 QY 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
 DB 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
 QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180
 QY 181 TWCTVVLONOKKVEFKIDIV-----PCPAPEPKSCDKHTTCEPLLGGPSVFL 227
 DB 181 TWCTVVLONOKKVEFKIDIVLAPEKCCVCEPCPCPAP-----VAGPSVFL 227
 QY 228 FPPKPKDTLMSRPEVTCVVVDVSHEDPEVKFMNVYDGVENNAKTKPREBOYNSTRY 287
 DB 228 FPPKPKDTLMSRPEVTCVVVDVSHEDPEVKFMNVYDGVENNAKTKPREBOYNSTRY 287
 QY 288 VSVLTVLHQMNLNGKEYKCVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ 347
 DB 288 VSVLTVLHQMNLNGKEYKCVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ 347
 QY 348 VSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSGSPFLYSKLTVDKSRMOQGNV 407
 DB 348 VSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSGSPFLYSKLTVDKSRMOQGNV 407
 QY 408 FSCSVMEALHNHYTQKSLSLSPG 431
 DB 408 FSCSVMEALHNHYTQKSLSLSPG 431

RESULT 14

ABG71122
 ID ABG71122 standard; protein; 432 AA.

XX ABG71122;
 AC
 XX 17-JAN-2003 (first entry)
 DT
 XX CD4-gamma2 chimeric heavy chain of the CD4-IgG2 chimeric protein.
 DE
 XX

```

KW CD4; gamma2 heavy chain; human immunodeficiency virus-1; HIV-1; mutant;
KM mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key
FT Peptide
FT Protein
FT
FT Location/Qualifiers
FT 1..25
FT /label= Signal_peptide
FT 26..432
FT /note= "Mature CD4-gamma2 chimeric heavy chain of the CD4
FT -IgG2 chimeric heterotetramer"
XX
XX US6451313-B1.
XX
XX 17-SEP-2002.
XX
XX 07-JUN-1995; 95US-00484681.
XX
XX 08-FEB-1991; 91US-00653684.
XX 10-FEB-1992; 92MO-US001143.
XX 08-DEC-1992; 92US-00960440.
XX
XX (PROG-) PROGENICS PHARM INC.
XX
XX Maddon PJ, Beaudry GA;
XX
XX WPI, 2003-038273/03.
XX DR N-PSDB; ABS55720.
XX
XX Novel CD4-immunoglobulin G2 chimeric heterotetramer neutralizes human
XX immunodeficiency virus-1 with two heavy and light chains encoded by
XX expression vectors designated CD4-IgG2HC-prcCMV and CD4-KLC-prcCMV,
XX respectively.
XX
XX Disclosure; Fig 3A-F; 54pp; English.
XX
XX The invention describes a purified CD4-immunoglobulin (Ig)G2 chimeric
XX heterotetramer (I) that neutralises human immunodeficiency virus-1 (HIV-
XX 1) having two heavy chains encoded by an expression vector designated CD4
XX -IgG2HC-prcCMV, and two light chains encoded by expression vector
XX designated CD4-KLC-prcCMV. (I) and a composition (II) comprising (I) or
XX (I) linked to a toxin, are useful for inhibiting HIV infection of a CD4
XX cell, and preventing a subject being infected with HIV by blocking the
XX spread of HIV infection. This is the amino acid sequence of the CD4-
XX gamma2 chimeric heavy chain of the CD4-IgG2 chimeric heterotetramer
XX useful in inhibiting HIV infection
XX
XX Sequence 432 AA:
XX
XX Query Match 87.9%; Score 2122; DB 6; Length 432;
XX Best Local Similarity 91.2%; Pred. No. 5,9e-112;
XX Matches 405; Conservative 8; Mismatches 5; Indels 26; Gaps 2;
XX
QY 1 MNRGVPRRLTLVLQALIPATOGKVVYLGKKDPTVELCTASQKSIQPHMKNNOIK 60
Db 1 MNRGVPRRLTLVLQALIPATOGKVVYLGKKDPTVELCTASQKSIQPHMKNNOIK 60
QY 61 ILGNQSPFLTKGPSKLNDRADSRSLMDGNFPLIIKNLKIDSDTYICEVDQKEEYQL 120
Db 61 ILGNQSPFLTKGPSKLNDRADSRSLMDGNFPLIIKNLKIDSDTYICEVDQKEEYQL 120
QY 121 LVFGILTANSDTHLLOGQSITLTLESPPGSSPSVQCSPPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGILTANSDTHLLOGQSITLTLESPPGSSPSVQCSPPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLNQGKQKVEFFIDIV-----PCPAPRPKSCDTHTCPELLGGSPVFL 227
Db 181 TWTCTVLNQGKQKVEFFIDIVLAPEKCCVCEPCPAP-----VAGPSVFL 227
QY 228 PPPKPDLTLMISRTPEVTCVVVDVSHEDPEVKFNNVYDDEVHNAKTKPRREQVNSTYRV 287
Db 228 PPPKPDLTLMISRTPEVTCVVVDVSHEDPEVKFNNVYDDEVHNAKTKPRREQVNSTYRV 287

```

```

QY 288 VSVLTVLHODWLNGKEYCKVSNKALPAPIEKTISKAKGQRPQOVYTLPPSRDELTKNQ 347
Db 288 VSVLTVLHODWLNGKEYCKVSNKGLPAPIEKTISKAKGQRPQOVYTLPPSRDELTKNQ 347
QY 348 VSLTCLVKGFPYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRNQQGNV 407
Db 348 VSLTCLVKGFPYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRNQQGNV 407
QY 408 FSCSYMEALHNHTYQKSLSLSPG 431
Db 408 FSCSYMEALHNHTYQKSLSLSPG 431

RESULT 15
AAR26783
ID AAR26783 standard; protein; 530 AA.
XX
XX AAR26783;
XX
XX AC 24-OCT-2003 (revised)
XX 25-MAR-2003 (revised)
XX DT 06-FEB-1993 (first entry)
XX
XX CD4-IgG2 chimeric heavy chain.
XX
XX homodimer; soluble CD4, T cell receptor; CD4 antigen; high recovery;
XX chimeric; increased serum half life; HIV infection; AIDS; ss.
XX
XX Homo sapiens.
XX Chimeric.
XX
XX Key Location/Qualifiers
XX Domain 1..205
XX /label= CD4 domain
XX Domain 206..302
XX /label= CH1 domain
XX Domain 303..312
XX /label= hinge domain
XX Domain 313..423
XX /label= CH2 domain
XX Domain 424..530
XX /label= CH3 domain
XX
XX WO9213947-A1.
XX
XX 20-AUG-1992.
XX
XX 10-FEB-1992; 92MO-US001143.
XX
XX 08-FEB-1991; 91US-00653684.
XX
XX (PROG-) PROGENICS PHARM INC.
XX
XX Beaudry GA, Maddon PJ;
XX
XX WPI, 1992-300034/36.
XX DR N-PSDB; AAQ28089.
XX
XX CD4-gamma-2 and CD4-IgG2 chimera(s) and expression vectors - for
XX treatment, prevention and diagnosis of HIV infection.
XX
XX Claim 15; Fig 4; 90pp; English.
XX
XX This sequence represents a CD4-IgG2 chimeric heavy chain heterotetramer
XX as described in AAQ28089 in Dhfr-CHO cells. The protein is efficiently
XX assembled intracellularly and effectively secreted from mammalian cells
XX pret. CHO, COS, or myeloma cells as a heterotetramer, enabling high
XX recovery and purification from the medium of cells expressing it. It
XX possesses increased serum half-life and has increased avidity for HIV cf.
XX heavy chain dimers. It can inhibit HIV infection of CD4+ cells and block
XX the spread of HIV infection within a patient. Attachment to a detectable

```

CC marker makes it useful in an assay for HIV or SIV infection, and it can
CC also be linked to toxins, eg Diphtheria, Pseudomonas exotoxin A (domains
CC I or II) or the deglycosylated A-chain of ricin. (Updated on 25-MAR-2003
CC to correct PN field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 530 AA;

Query Match 86.4%; Score 2085; DB 2; Length 530;
Best Local Similarity 77.3%; Pred. No. 9e-110;
Matches 409; Conservative 9; Mismatches 13; Indels 98; Gaps 4;

QY 1 MNRGVPFRLHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
DB 1 MNRGVPFRLHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPPLIKNLKIEDSDTYICEVEDQKEEVQL 120
DB 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPPLIKNLKIEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSPTHLLQGGSLTLTLSPSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGGSLTLTLSPSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTCTVLQGNKKVEFKIDIV-----PCPA-----PEP 207
DB 181 TWCTCTVLQGNKKVEFKIDIVLAFASATGKPSVFLAPCSRSTSESTALGCLVNDYFPEP 240
QY 208 ----- 207
DB 241 VTVSNMNGALTSGVHTFPAVLQSSGLYSLSVTVPSNFGTQTYTCNVDHKPSNTKYDK 300
QY 208 ----KSCDKHTCP-ELLGSPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNW 262
DB 301 TVERKCCVCEPCPAPVAGPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNW 360
QY 263 YVDGVEVNAKTKPREQYNSTYRVSVLTVLDHQMNGEKYCKVSKNKPAPLEKTI 322
DB 361 YVDGVEVNAKTKPREQYNSTYRVSVLTVLDHQMNGEKYCKVSKNKPAPLEKTI 420
QY 323 KAKGQPREPQVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVMESNGQPENNYKTPPV 382
DB 421 KTKGQPREPQVYTLPPSRDEEMTKNOVSLTCLVKGFPYSDIAVMESNGQPENNYKTPPV 480
QY 383 LDSGSPFLYSGKLTVDKSRMOQGNVFCSCVMHEALHNHYTQKSLSLSPG 431
DB 481 LDSGSPFLYSGKLTVDKSRMOQGNVFCSCVMHEALHNHYTQKSLSLSPG 529

RESULT 16

AA85080
ID AA85080 standard; protein; 530 AA.

XX AA85080;

XX 19-JUN-2000 (first entry)

DE CD4-IgG2 chimeric heterotetramer heavy chain amino acid sequence.

XX CD4-IgG2 chimeric heavy chain heterotetramer; immunoconjugate; treatment;
XX cytotoxic radionuclide; cell surface glycoprotein; prevent; infection;
XX cellular immune response interaction mediator; HIV interaction; staging;
XX prognosis; envelope glycoprotein burden; human.

XX Homo sapiens.

XX OS

XX US6034223-A.

XX 07-MAR-2000.

XX 07-JUN-1995; 95US-00477460.

XX 07-AUG-1992; 92US-00927931.

XX 06-AUG-1993; 93WO-US0074422.

PR 03-FEB-1995; 95US-00379516.

XX (PROG-) PROGENICS PHARM INC.

XX Allaway GP, Maddon PJ;

XX WPI; 2000-269502/23.

DR N-PSDB; AA298856.

PT New immunoconjugate, used to treat, prevent or image human immune

PT deficiency virus infection, comprises radionuclide attached to

PT heterotetramer of CD4-immunoglobulin chimeras.

PS Disclosure; Fig 4; 58pp; English.

CC This sequence represents the CD4-IgG2 chimeric heavy chain amino acid
CC sequence from the CD4-IgG2 chimeric heterotetramer. The invention relates
CC to an immunoconjugate comprising a cytotoxic radionuclide and a
CC heterotetramer of two heavy chains and two light chains. The cytotoxic
CC radionuclide is linked to either the heavy chains or the light chains, or
CC to all four chains, directly or through a bifunctional chelator. Both
CC heavy chains are chimeric CD4-Ig (immunoglobulin) G2 chains encoded by
CC vector CD4-IgG2HC-PRCCMV (ATCC 75193) and both light chains are chimeric
CC CD4-keappa chains encoded by vector CD4-KLC-PRCCMV (ATCC 75194). CD4 is a
CC non-polymorphic cell surface glycoprotein that is expressed on the
CC surface of helper T lymphocytes, cells of the monocyte/macrophage lineage
CC and dendritic cells. CD4 associates with major histocompatibility complex
CC (MHC) class II molecules on the surface of antigen presenting cells to
CC mediate efficient cellular immune response interactions. In humans CD4 is
CC the target of interaction with the human immunodeficiency virus HIV. The
CC immunoconjugate is used to kill cells infected with HIV, and for treating
CC or preventing infection. It is also used for imaging HIV-infected tissues
CC (for staging or prognosis of infection, and for assessing efficacy of
CC treatments). The immunoconjugate is also used to determine the HIV
CC envelope glycoprotein burden, once determined, this information is used
CC in the staging and prognosis of HIV infected patients. The
CC immunoconjugate should be active against all strains of HIV (since the
CC CD4-gp120 interaction is essential for infection). The heterotetramers
CC are assembled intracellularly and secreted efficiently from mammalian
CC cells, allowing high recovery and purification from the culture medium.
CC They have longer half-life in serum and greater avidity than heavy chain
CC dimers

XX Sequence 530 AA;

Query Match 86.4%; Score 2085; DB 3; Length 530;
Best Local Similarity 77.3%; Pred. No. 9e-110;
Matches 409; Conservative 9; Mismatches 13; Indels 98; Gaps 4;

QY 1 MNRGVPFRLHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
DB 1 MNRGVPFRLHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60
QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPPLIKNLKIEDSDTYICEVEDQKEEVQL 120
DB 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPPLIKNLKIEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSPTHLLQGGSLTLTLSPSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGGSLTLTLSPSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTCTVLQGNKKVEFKIDIV-----PCPA-----PEP 207
DB 181 TWCTCTVLQGNKKVEFKIDIVLAFASATGKPSVFLAPCSRSTSESTALGCLVNDYFPEP 240
QY 208 ----- 207
DB 241 VTVSNMNGALTSGVHTFPAVLQSSGLYSLSVTVPSNFGTQTYTCNVDHKPSNTKYDK 300
QY 208 ----KSCDKHTCP-ELLGSPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNW 262
DB 301 TVERKCCVCEPCPAPVAGPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNW 360

```

Oy 263 YVDGEVHNAKTKPREBOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKTIS 322
Db 361 YVDGEVHNAKTKPREBOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKTIS 420
Oy 333 KAKGPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPV 382
Db 421 KTKGPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPM 480
Oy 383 LDSGSPFLYSKLTVDKSRMGOGNVFCSSVMHEALHNHYTQKSLSLSPG 431
Db 481 LDSGSPFLYSKLTVDKSRMGOGNVFCSSVMHEALHNHYTQKSLSLSPG 529

RESULT 17
AAB67323
ID AAB67323 standard; protein; 530 AA.
XX
XX AAB67323;
AC
XX
XX 23-APR-2001 (first entry)
DT
XX CD4-IgG2 chimeric heavy chain protein.
DE
XX Immunoconjugate; chelator; chimeric; HIV; human immunodeficiency virus.
KM
XX Homo sapiens.
OS
XX US6177549-B1.
PN
XX 23-JAN-2001.
PD
XX 10-JUN-1999; 99US-00329916.
PF
XX
XX 07-AUG-1992; 92US-00927931.
PR
XX 06-AUG-1993; 93WO-US007422.
PR
XX 03-FEB-1995; 95US-00379516.
PR
XX 07-JUN-1995; 95US-00477460.
XX
XX (PROG-) PROGENICS PHARM INC.
PA
XX
XX Maddon PJ, Allaway GP;
PI
XX
XX WPI; 2001-158582/16.
DR
XX
XX Immunoconjugate for treating human immunodeficiency virus-infected
PT subject, consists of cytotoxic radionuclide linked to heterotetramer
PT comprising two chimeric CD4-IgG2 heavy chains and two chimeric CD4-
PT kappa1 light chains.
XX
XX Disclosure; Fig 4; 43pp; English.
PS
XX
XX The present invention relates to an immunoconjugate, comprising a
CC cytotoxic radionuclide linked, directly or via a bifunctional chelator,
CC to a heterotetramer having two chimeric CD4-IgG2 heavy chains encoded by
CC an expression vector CD4-IgG2HC-pRCWV and two chimeric CD4-kappa 1 light
CC chains encoded by an expression vector CD4-KLC-pRCWV. The invention is
CC useful for killing human immunodeficiency virus (HIV)-infected cells, for
CC the treatment and prevention of infection with HIV
XX
XX Sequence 530 AA;
SQ

```

```

Query Match 86.4%; Score 2085; DB 4; Length 530;
Best Local Similarity 77.3%; Pred. No. 9e-110;
Matches 409; Conservative 9; Mismatches 13; Indels 98; Gaps 4;

```

```

Oy 1 MKRGVPRRLTLVLQALPAATOGNKVYLGGKGDVLTCTASQKSIQFMKNSNOIK 60
Db 1 MKRGVPRRLTLVLQALPAATOGNKVYLGGKGDVLTCTASQKSIQFMKNSNOIK 60
Oy 61 IIVGNQSFLLTKGPKSKLNDRAISRSLMDQGNFPLIIKLIKIDSDTYICEVEDQKEVQL 120
Db 61 IIVGNQSFLLTKGPKSKLNDRAISRSLMDQGNFPLIIKLIKIDSDTYICEVEDQKEVQL 120

```

```

Oy 121 LVFGLTANSDTHLLOQSLTTLTLESPGSSPSVOCRSRPGKNIQGGKTLISVQLEODSG 180
Db 121 LVFGLTANSDTHLLOQSLTTLTLESPGSSPSVOCRSRPGKNIQGGKTLISVQLEODSG 180
Oy 181 TWCTVLQONQKVEFKIDIV-----PCPA-----PEP 207
Db 181 TWCTVLQONQKVEFKIDIVVLAFASFTGSPVFLAPCSRSTSBTALGCLVKDYFPEP 240
Oy 208 ----- 207
Db 241 VTVMNSGALTSVHTPEPAVLQSSGLYSLSVTVTPSSNFGTQTYTCNVDHKPSNTKYDK 300
Oy 208 ----KSCDKHTICP-ELLGSPVFLFPKPKDITMISRTPEVTCVNVDSHEDPEVKFN 262
Db 301 TVRKCCVCEPCPCAPPAAGSVFLFPKPKDITMISRTPEVTCVNVDSHEDPEVKFN 360
Oy 263 YVDGEVHNAKTKPREBOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKTIS 322
Db 361 YVDGEVHNAKTKPREBOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKTIS 420
Oy 323 KAKGPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPV 382
Db 421 KTKGPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPM 480
Oy 383 LDSGSPFLYSKLTVDKSRMGOGNVFCSSVMHEALHNHYTQKSLSLSPG 431
Db 481 LDSGSPFLYSKLTVDKSRMGOGNVFCSSVMHEALHNHYTQKSLSLSPG 529

RESULT 18
AAB80884
ID AAB80884 standard; protein; 530 AA.
XX
XX AAB80884;
AC
XX
XX 29-MAY-2001 (first entry)
DT
XX
XX Human CD4-IgG2 chimeric heavy chain.
DE
XX
XX Human; Anti-HIV; CD4-IgG2 chimeric heterotetramer;
KM immunoglobulin gamma 2.
XX
XX Homo sapiens.
OS
XX
XX US6187748-B1.
PN
XX 13-FEB-2001.
PD
XX
XX 07-JUN-1995; 95US-00485372.
PF
XX
XX 08-FEB-1991; 91US-00553684.
PR
XX 10-FEB-1992; 92WO-US001143.
PR
XX 08-DEC-1992; 92US-00960440.
XX
XX (PROG-) PROGENICS PHARM INC.
PA
XX
XX Maddon PJ, Beaudry GA;
PI
XX
XX WPI; 2001-264981/27.
DR
XX
XX N-PSDB; AAF77830.
DT
XX
XX Inhibiting human immunodeficiency virus (HIV) infection of a CD4+ cell,
PT or treating a subject having CD4+ cells infected with HIV involves using
PT CD4-IgG2 chimeric heterotetramer to form a complex with the HIV.
XX
XX Disclosure; Fig 4; 55pp; English.
PS
XX
XX The present invention relates to a method for inhibiting infection of a
CC CD4+ cell by HIV. The method comprises contacting the HIV with a CD4-IgG2
CC chimeric heterotetramer to form a complex with the HIV (CD4 = cluster of
CC differentiation 4; IgG2 = immunoglobulin gamma 2). CD4 is a cell surface
CC glycoprotein that is expressed primarily on the surface of T cells. In
CC man, CD4 is the target of interaction with HIV. The heterotetramer has

```

CC two heavy and two light chains which are encoded by expression vectors
 CC CD4-IgG2HC-PRCCMV (Y1) and CD4-KLC-PRCCMV (Y2), respectively. The method
 CC is used to inhibit infection of a CD4+ cell by a HIV and to prevent CD4+
 CC cells of a subject from becoming infected with HIV. The method is also
 CC useful for treating a subject having CD4+ cells infected with HIV. The
 CC present sequence is human fusion protein: CD4-Ig2 chimeric heavy chain of
 CC the CD4-IgG2 chimeric heterotetramer. This sequence was used in the
 CC method of the present invention

XX Sequence 530 AA;

Query Match 86.4%; Score 2085; DB 4; Length 530;
 Best Local Similarity 77.3%; Pred. No. 9e-110;
 Matches 409; Conservative 9; Mismatches 13; Indels 98; Gaps 4;

```

QY 1 MNRGVPFRHLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPFRHLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSPTHLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLQNKVKVEFKIDIV-----PCPA-----PEP 207
DB 181 TWTCVTLQNKVKVEFKIDIVLAFASTGKPSVFLPAPCSRSTSESTALGCLVMDYFPEP 240
QY 208 ----- 207
DB 241 VTVMNSGALTSQVHTFPFAVLQSSGLYSLSVVTVPSSNFGQTGYTCNVHDKPSNTKYDK 300
QY 208 ----KSCDKHTHCP-ELLGGPSVFLPPPKPKDTLMTSRPEVTCVVDVSHEDPEVKNW 262
DB 301 TVERKCCVCEPCPCAPAPVAGPSVFLPPPKPKDTLMTSRPEVTCVVDVSHEDPEVQFNW 360
QY 263 YVDGVEVNAKTKPREBOYNSTFRVSVTLVHODMLNGEKYKCVSNKGLPAPIEKTIS 322
DB 361 YVDGVEVNAKTKPREBOYNSTFRVSVTLVHODMLNGEKYKCVSNKGLPAPIEKTIS 420
QY 323 KAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPV 382
DB 421 KTKGQPREPOVYTLPPSRDEMTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPM 480
QY 383 LDSGSEFLYLSKLTVDKSRMOQGNVFCSSVMHEALHNHYTOKSLSLSPG 431
DB 481 LDSGSEFLYLSKLTVDKSRMOQGNVFCSSVMHEALHNHYTOKSLSLSPG 529

RESULT 19
ABG71123
ID ABG71123 standard; protein; 530 AA.
AC ABG71123;
XX 17-JAN-2003 (first entry)
DE CD4-immunoglobulin G2 (IgG2) chimeric heterotetramer.
XX CD4; immunoglobulin G2; Ig gamma2; human immunodeficiency virus-1; HIV-1;
XX mutant; muteln.
XX Homo sapiens.
XX Synthetic.
XX Key Location/Qualifiers
XX Peptide 1..25
XX Protein /label= Signal_peptide
XX 26..530
XX /note= "Mature CD4-IgG2 chimeric heterotetramer"

```

```

XX US6451313-B1.
XX 17-SEP-2002.
XX 07-JUN-1995; 95US-00484681.
XX 08-FEB-1991; 91US-00653684.
XX 10-FEB-1992; 92WO-US001143.
XX 08-DEC-1992; 92US-00960440.
XX (PROG-) PROGENICS PHARM INC.
XX Maddon PJ, Beaudry GA;
XX WPI: 2003-038273/03.
XX N-PDB; ABS5721.
XX Novel CD4-immunoglobulin G2 chimeric heterotetramer neutralizes human
XX immunodeficiency virus-1 with two heavy and light chains encoded by
XX expression vectors designated CD4-IgG2HC-PRCCMV and CD4-KLC-PRCCMV,
XX respectively.
XX Claim 1; Fig. 4A-H; 54pg; English.
XX The invention describes a purified CD4-immunoglobulin (Ig)G2 chimeric
XX heterotetramer (I) that neutralises human immunodeficiency virus-1 (HIV-
XX 1) having two heavy chains encoded by an expression vector designated CD4
XX -IgG2HC-PRCCMV, and two light chains encoded by expression vector
XX designated CD4-KLC-PRCCMV. (I) and a composition (II) comprising (I) or
XX (I) linked to a toxin, are useful for inhibiting HIV infection of a CD4
XX cell, and preventing a subject being infected with HIV by blocking the
XX spread of HIV infection. This is the amino acid sequence of the CD4-
XX immunoglobulin G2 chimeric heavy chain chimeric heterotetramer useful in
XX inhibiting HIV infection
XX Sequence 530 AA;
XX
Query Match 86.4%; Score 2085; DB 6; Length 530;
Best Local Similarity 77.3%; Pred. No. 9e-110;
Matches 409; Conservative 9; Mismatches 13; Indels 98; Gaps 4;
QY 1 MNRGVPFRHLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPFRHLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSPTHLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLQNKVKVEFKIDIV-----PCPA-----PEP 207
DB 181 TWTCVTLQNKVKVEFKIDIVLAFASTGKPSVFLPAPCSRSTSESTALGCLVMDYFPEP 240
QY 208 ----- 207
DB 241 VTVMNSGALTSQVHTFPFAVLQSSGLYSLSVVTVPSSNFGQTGYTCNVHDKPSNTKYDK 300
QY 208 ----KSCDKHTHCP-ELLGGPSVFLPPPKPKDTLMTSRPEVTCVVDVSHEDPEVKNW 262
DB 301 TVERKCCVCEPCPCAPAPVAGPSVFLPPPKPKDTLMTSRPEVTCVVDVSHEDPEVQFNW 360
QY 263 YVDGVEVNAKTKPREBOYNSTFRVSVTLVHODMLNGEKYKCVSNKGLPAPIEKTIS 322
DB 361 YVDGVEVNAKTKPREBOYNSTFRVSVTLVHODMLNGEKYKCVSNKGLPAPIEKTIS 420
QY 323 KAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPV 382
DB 421 KTKGQPREPOVYTLPPSRDEMTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPM 480

```

```
QY 363 LDSDSGFLYSLTVDKSRMOQGNVFCGVMEALHNHYTOKSLSPG 431
DB 481 LDSDSGFLYSLTVDKSRMOQGNVFCGVMEALHNHYTOKSLSPG 529

RESULT 20
ID AAY51082
AA51082 standard; protein; 616 AA.
AC AAY51082;
XX
XX 23-MAR-2000 (first entry)
DT
XX
XX Human fusion protein CD4Bgammal.
DE
XX
XX Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
KW anti-human immunodeficiency virus; CD4Bgammal.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX US6004781-A.
XX
XX 21-DEC-1999.
XX
XX 04-FEB-1994; 94US-00191708.
XX
XX 22-JAN-1988; 88US-00147351.
XX 23-JAN-1989; 89US-0029596.
XX 09-JUN-1992; 92US-00896781.
XX 12-APR-1993; 93US-00057952.
XX
XX (GEO) GEN HOSPITAL CORP.
XX
XX Seed B;
XX
XX MPI: 2000-085792/07.
XX
XX N-PSDB; AA244065.
XX
XX Fusion protein useful for the treatment of human immunodeficiency virus.
XX
XX Example 1; Col 59-70; 39pp; English.
XX
XX This invention describes a novel nucleic acid (I) encoding a fusion
XX protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
XX and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
XX chain (III). The products of the invention have anti-human
XX immunodeficiency virus (HIV) activity and are capable of binding to
XX gp120. The fusion protein is useful for treating human immunodeficiency
XX virus (HIV) or simian immunodeficiency virus (SIV). This sequence
XX represents the fusion protein CD4Bgammal which is constructed from CD4
XX linked to human IgG1 upstream of the hinge region
XX
XX Sequence 616 AA;

Query Match 86.2%; Score 2081; DB 3; Length 616;
Best Local Similarity 68.1%; Pred. No. 1.8e-109;
Matches 419; Conservative 1; Mismatches 11; Indels 184; Gaps 3;

QY 1 MNRGVPFRHLVLVLTALPAATQGNKVVLGKGGDTVELTCTASQKKSIOFMKNSNOIK 60
DB 1 MNRGVPFRHLVLVLTALPAATQGNKVVLGKGGDTVELTCTASQKKSIOFMKNSNOIK 60
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPPLIIRKLIKEDSDTYICEVEDQKEEYQL 120
DB 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPPLIIRKLIKEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGITANSDBTHLQGSQITLTLESPPGSSSVQCGSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGITANSDBTHLQGSQITLTLESPPGSSSVQCGSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGITANSDBTHLQGSQITLTLESPPGSSSVQCGSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGITANSDBTHLQGSQITLTLESPPGSSSVQCGSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLTQNKQKVEFKDIV----- 200
```

```
DB 181 TWTCTVLTQNKQKVEFKDIVLTALFQKASSIYVKKEGQVERSPPLAFTVEXTLQSGELMW 240
QY 201 -----PCPAPE----- 206
DB 241 QAERASSKSWITFDLKNKEVSVKRVTDPRKLQMGKLLPLHLTLFQALPQYAGSGLTLA 300
QY 207 -----PKSCDKTHNC----- 216
DB 301 LEAKTGKLHQEVNLVYMRATOLQKNLTCEVWGPTSPKMLSLKLNKEAKVSKREKPYWV 360
QY 217 -----DELIGGPEVFLFPPEKDTL 236
DB 361 LNPBAGMWQCLSDSGVLESNIKVLPWTSTPVHADPEAPDELIGGPEVFLFPPEKDTL 420
QY 237 MISRTPEVTCVVDVSHEDPEVKFNWYDGYEVHNAKTKPREBOYSTYRVVSVLTVLHQ 296
DB 421 MISRTPEVTCVVDVSHEDPEVKFNWYDGYEVHNAKTKPREBOYSTYRVVSVLTVLHQ 480
QY 297 DMLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKGNVSLTCLVKG 356
DB 481 DMLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKGNVSLTCLVKG 540
QY 357 FYPSDIAVWESNQGPENNYKTTTPVLDSDGSFFLYSLTVDKSRMOQGNVFCGVMEHA 416
DB 541 FYPSDIAVWESNQGPENNYKTTTPVLDSDGSFFLYSLTVDKSRMOQGNVFCGVMEHA 600
QY 417 LHNHYTOKSLSPG 431
DB 601 LHNHYTOKSLSPG 615

RESULT 21
AAY59172
ID AAY59172 standard; protein; 616 AA.
XX
XX AAY59172;
XX
XX 14-MAR-2000 (first entry)
XX
XX CD4-Ig fusion protein CD4Bgammal.
XX
XX HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
XX secreted protein; SIV infection; medicament.
XX
XX Synthetic.
XX
XX Homo sapiens.
XX
XX CA1340741-C.
XX
XX 14-SEP-1999.
XX
XX 20-JAN-1989; 89CA-00588749.
XX
XX 20-JAN-1989; 89CA-00588749.
XX
XX 20-JAN-1989; 89CA-00588749.
XX
XX (GEO) GEN HOSPITAL CORP.
XX
XX Seed B;
XX
XX MPI: 2000-063015/06.
XX
XX N-PSDB; AA248205.
XX
XX New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
XX the treatment of HIV or simian immunodeficiency virus infections.
XX
XX Example 1; Page 61-68; 89pp; English.
XX
XX The invention provides a fusion gene encoding a fusion protein that
XX comprises an extracellular CD4 DNA sequence or its fragment which binds
XX to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
XX sequence of an Ig heavy or light chain, where the DNA sequence encoding
XX the variable region has been replaced with the DNA sequence which encodes
```

CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
 CC capable of being secreted. The fusion proteins are useful for treating
 CC HIV or SIV infections in animals, preferably humans. They are also useful
 CC for producing medicaments which can be used for treating HIV or SIV
 CC infections in humans. The present sequence represents the fusion protein
 CC CD4gamma1 where the CD4 is linked to human IgG1 at the BamI site
 CC downstream from the hinge region

XX Sequence 616 AA;

Query Match 86.2%; Score 2081; DB 3; Length 616;
 Best Local Similarity 68.1%; Pred. No. 1,8e-109;
 Matches 419; Conservative 1; Mismatches 11; Indels 184; Gaps 3;

QY 1 MNRGVPFRHLILVLTALPAATQGNKVLGKKGDVETLTCTASQKKSIOFHMKNNOIK 60
 DB 1 MNRGVPFRHLILVLTALPAATQGNKVLGKKGDVETLTCTASQKKSIOFHMKNNOIK 60
 QY 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 DB 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 QY 121 LVFGITANSPTHTLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGITANSPTHTLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTVLQNOKKVEFKIDIV----- 200
 DB 181 TWCTVLQNOKKVEFKIDIVLAFOKASSIVYKKEGQVRSFPLAFVEKLTGSGELMW 240
 QY 201 -----PCPAPE----- 206
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKCLPLHLTLPOALPOYAGSGNLTLA 300
 QY 207 -----PKSCDKTHTC----- 216
 DB 301 LEAKTGKLGQEVNLVVMRATQLOKNLTCEVWGPTSPKLMLSLKENKEAKVSKREKPVW 360
 QY 217 -----PELGGPSVPLFPKPKDYL 236
 DB 361 LNPEAGMQCLSDSGQVLEESNIVLPTWSTPVHADPEABELGGPSVPLFPKPKDYL 420
 QY 237 MISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVHNAKTKPREQYNSYTRVSVLTVLHQ 296
 DB 421 MISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVHNAKTKPREQYNSYTRVSVLTVLHQ 480
 QY 297 DMLNKEKCKVSNKALPAPLEKTKSKAGQREPOVYTLPPSRDELTKNOVSLTCLVKG 356
 DB 481 DMLNKEKCKVSNKALPAPLEKTKSKAGQREPOVYTLPPSRDELTKNOVSLTCLVKG 540
 QY 357 FYPSPDIAVEMESNGQPENNYKTPVLDSDGSFPLYSKLTVDKSRMQGNVFSQVMHEA 416
 DB 541 FYPSPDIAVEMESNGQPENNYKTPVLDSDGSFPLYSKLTVDKSRMQGNVFSQVMHEA 600
 QY 417 LHNHYTQKSLSPG 431
 DB 601 LHNHYTQKSLSPG 615

RESULT 22
 AAP93008
 ID AAP93008 standard; protein; 729 AA.
 XX AAP93008;
 AC
 XX 25-MAR-2003 (revised)
 DT 02-NOV-1992 (first entry)
 XX
 XX Genetic construct which encodes CD4 linked to human IgG1 at the Hind3
 DE site upstream of the CH1 region (fusion protein CD4H-gamma-1).
 XX Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
 KW diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.

XX Homo sapiens.
 OS
 XX EP25262-A.
 PN
 XX 26-JUL-1989.
 PD
 XX 20-JAN-1989; 89EP-00100913.
 PF
 XX 22-JAN-1988; 88US-00147351.
 PR
 XX (GENO) GEN HOSPITAL CORP.
 PA
 XX Seed B;
 PI
 XX WPI: 1989-214472/30.
 DR N-PSDB; AAN90356.

Example; Table 1, Page 12-23; 68pp; English.

CC The fusion protein genes of the invention pref. comprise cDNA sequences
 CC which encode CD4 or a fragment which binds gp120 ligated to an expression
 CC plasmid which encodes an antibody in which the variable region of the
 CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion
 CC protein may comprise the complete CD4 sequence, the 370 AA extracellular
 CC region and the membrane spanning domain, or the extracellular region. The
 CC Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
 CC specifically claimed: fusion proteins CD4H-gamma-1, CD4Mu, CD4Pmu, CD4E-
 CC gamma1, and CD4Mu (No. 67608), pCD4P-gamma (No. 67609) and pCD4E-gamma-1
 CC (No. 67610). The plasmid containing (pCD4H-gamma-1) has been deposited in
 CC E. coli (MC1061/B3) at the ATCC under accession number 67611. (Updated on
 CC 25-MAR-2003 to correct PA field.)

XX Sequence 729 AA;

Query Match 86.1%; Score 2079.5; DB 1; Length 729;
 Best Local Similarity 58.8%; Pred. No. 2,6e-109;
 Matches 428; Conservative 0; Mismatches 3; Indels 297; Gaps 4;

QY 1 MNRGVPFRHLILVLTALPAATQGNKVLGKKGDVETLTCTASQKKSIOFHMKNNOIK 60
 DB 1 MNRGVPFRHLILVLTALPAATQGNKVLGKKGDVETLTCTASQKKSIOFHMKNNOIK 60
 QY 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 DB 61 ILNGGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120
 QY 121 LVFGITANSPTHTLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGITANSPTHTLLQGQSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWCTVLQNOKKVEFKIDIV----- 200
 DB 181 TWCTVLQNOKKVEFKIDIVLAFOKASSIVYKKEGQVRSFPLAFVEKLTGSGELMW 240
 QY 201 -----PC----- 202
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKCLPLHLTLPOALPOYAGSGNLTLA 300
 QY 201 -----PC----- 200
 DB 301 LEAKTGKLGQEVNLVVMRATQLOKNLTCEVWGPTSPKLMLSLKENKEAKVSKREKPVW 360
 QY 201 -----PC----- 202
 DB 361 LNPEAGMQCLSDSGQVLEESNIVLPTWSTPVHADPEASTKQSPVPLAPSSKSTSGG 420
 QY 203 -----PAP----- 205
 DB 421 TAAAGCLVKDYFPEPPTVSNMGSALTSVHTFPAVLQSGLSGLYSVTVTPSSSLGTQTY 480

```

QY 206 -----EPKSCDKHTTC-----PELLGGPSVFLPPPKDITLMISTRPE 243
DB 481 ICNVNHNKPSNTKVDKKNVEPKSCDKHTTCPPCPAPPELLGGPSVFLPPPKDITLMISTRPE 540
QY 244 VTCVVVDVSHEDPEVKFNNYVVDGVEVHNAKTKRREQDYNSTRYVSVLTVLHQMVLNGKE 303
DB 541 VTCVVVDVSHEDPEVKFNNYVVDGVEVHNAKTKRREQDYNSTRYVSVLTVLHQMVLNGKE 600
QY 304 YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELITNQVSLTCLVKGFTPSDIA 363
DB 601 YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELITNQVSLTCLVKGFTPSDIA 660
QY 364 VEMESNGQPENNYKTTTPVLDSGSPFLYSKLTVDKSRMQQGNVSCSVMEHALLNHYTQ 423
DB 661 VEMESNGQPENNYKTTTPVLDSGSPFLYSKLTVDKSRMQQGNVSCSVMEHALLNHYTQ 720
QY 424 KSLSLSPG 431
DB 721 KSLSLSPG 728

RESULT 23
AAB19507
ID AAB19507 standard; protein; 729 AA.
XX
AC AAB19507;
XX
DT 09-JAN-2001 (first entry)
XX
DE CD4-IgG1 fusion protein CH4Hgamma1.
XX
KM CD4; IgG1; human; CD4Hgamma1; fusion protein; immunoglobulin; HIV; SIV;
XX
OS gp120; therapy; diagnosis.
XX
OS Homo sapiens.
XX
FH Key 1.395 Location/Qualifiers
FT Protein /note="CD4 extracellular region"
FT Protein 400..729
FT Protein /note="IgG1 heavy chain"
XX
PN US6117656-A.
XX
PD 12-SEP-2000.
XX
PF 07-JUN-1995; 95US-00479353.
XX
PR 23-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00299596.
PR 09-JUN-1992; 92US-00896781.
PR 12-APR-1993; 93US-00057952.
PR 04-FEB-1994; 94US-00191708.
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI: 2000-586558/55.
DR N-PSDB; AAB50660.
XX
FT CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
FT SIV.
XX
PS Example 1; Col 13-30; 39pp; English.
XX
CC The present sequence is that of fusion protein CD4Hgamma1 comprising the
CC extracellular portion of CD4, which binds to HIV gp120, linked at its C-
CC terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
CC encoding CD4 was linked to IgG1 DNA at the HindIII site upstream of the CH1
CC region (see AAB50660). Fusion protein CD4Hgamma1 and a nucleic acid
CC encoding it are claimed. Also claimed are a vector comprising the nucleic

```

```

CC acid, and a method of producing the fusion protein in secreted form using
CC a transformed host cell. The fusion protein may further comprise a
CC therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
CC can be administered to an animal (including humans) for treatment of HIV
CC or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC and tissue staining. IgG1 fusion proteins such as CD4Hgamma1 provide both
CC complement-mediated and cell-mediated immunity
SQ Sequence 729 AA:
Query Match 86.1%; Score 2079.5; DB 3; Length 729;
Best Local Similarity 58.8%; Pred. No. 2.6e-109;
Matches 428; Conservative 0; Mismatches 3; Indels 297; Gaps 4;
QY 1 NMRGVPFRHLVLVLDLALPPATQGNKYLGGKGTVELTCTASQKSIQFMWKNQIK 60
DB 1 NMRGVPFRHLVLVLDLALPPATQGNKYLGGKGTVELTCTASQKSIQFMWKNQIK 60
QY 61 ILNGGSPFLTKGSPKLNDRADSRSLMDQGNFLLIKULKIDSDTYICEVEDQKEVQL 120
DB 61 ILNGGSPFLTKGSPKLNDRADSRSLMDQGNFLLIKULKIDSDTYICEVEDQKEVQL 120
QY 121 LVFGILTANSDTHLQGGSLTTLTSSPPGSSPVQCSRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLQGGSLTTLTSSPPGSSPVQCSRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKKEKIDIV----- 200
DB 181 TWTCTVLQNKKEKIDIVVLAFOKASSIVYKKEGEVEFSPPLAFVEXLTGSGELMW 240
QY 201 ----- 200
DB 241 QAEKSSSKSWITFDLKNKEVSVKAVTODPKLQMGKPLHLTLPQALPQYAGSGLTLA 300
QY 201 ----- 200
DB 301 LEAKTGKLEQEVNLYVMEATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVV 360
QY 201 -----PC----- 202
DB 361 INPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVHADPEASTKGPSVPLAPSKSTSGG 420
QY 203 -----PAP----- 205
DB 421 TTAALGLVSYRPEPVTVSNNSGALTSGVNTPPVALQSSGLYSLSVVTVBSSISGTQTY 480
QY 206 -----EPKSCDKHTTC-----PELLGGPSVFLPPPKDITLMISTRPE 243
DB 481 ICNVNHNKPSNTKVDKKNVEPKSCDKHTTCPPCPAPPELLGGPSVFLPPPKDITLMISTRPE 540
QY 244 VTCVVVDVSHEDPEVKFNNYVVDGVEVHNAKTKRREQDYNSTRYVSVLTVLHQMVLNGKE 303
DB 541 VTCVVVDVSHEDPEVKFNNYVVDGVEVHNAKTKRREQDYNSTRYVSVLTVLHQMVLNGKE 600
QY 304 YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELITNQVSLTCLVKGFTPSDIA 363
DB 601 YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELITNQVSLTCLVKGFTPSDIA 660
QY 364 VEMESNGQPENNYKTTTPVLDSGSPFLYSKLTVDKSRMQQGNVSCSVMEHALLNHYTQ 423
DB 661 VEMESNGQPENNYKTTTPVLDSGSPFLYSKLTVDKSRMQQGNVSCSVMEHALLNHYTQ 720
QY 424 KSLSLSPG 431
DB 721 KSLSLSPG 728

RESULT 24
AAY59168
ID AAY59168 standard; protein; 729 AA.
XX
AC AAY59168;
XX

```


PT Non-peptidyl toxin or radionuclide and CD4-gamma 2 or CD4-IgG2
PT Immunoconjugates - used to kill HIV-infected cells and to image and
PT stage HIV infection.

XX Disclosure; Fig 4, 142pp; English.

XX A tetramer comprising 2 IgG2 heavy chains or two CD4-IgG2 chimeric heavy
XX chains and two kappa light chains or CD4-kappa light chains (AAR46680)
XX linked to a non-peptidyl toxin or a gamma radiation-emitting radionuclide
XX of low to moderate cytotoxicity. The resulting immunoconjugate comprising
XX the toxin can be used to kill HIV infected cells and to treat HIV
XX infected subjects to reduce the population of HIV infected cells. It can
XX also be used to reduce the likelihood of infection. The immunoconjugate
XX comprising the radionuclide can be used to image HIV infected tissue, to
XX calculate the stage of HIV infection or the efficacy of an anti-HIV
XX treatment using the imaging technique and for determining the prognosis
XX of an HIV infected subject. (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 530 AA;

Query Match 86.0%; Score 2075; DB 2; Length 530;
Best Local Similarity 76.7%; Pred. No. 3.3e-109;
Matches 406; Conservative 12; Mismatches 13; Indels 98; Gaps 4;

QY 1 MNRGVPFRLHLVQLALPAATQGNKVVLGKGGDTVELCTASQKSIQFMKNSNOIK 60
DB 1 MNRGVPFRLHLVQLALPAATQGNKVVLGKGGDTVELCTASQKSIQFMKNSNOIK 60
QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVQL 120
DB 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGILTANSDDTHLLOQGSITLTLESPPGSSBVOCSPRGKNIQGGKTLVSQLELDPSG 180
DB 121 LVFGILTANSDDTHLLOQGSITLTLESPPGSSBVOCSPRGKNIQGGKTLVSQLELDPSG 180
QY 181 TWTCTVLONOKKVEFKIDIV-----PCPA-----PEP 207
DB 181 TWTCTVLONOKKVEFKIDIVLAAPASTKGPSPFLAPCSRSTSESTALGCLVKDYFPEP 240
QY 208 ----- 207
DB 241 VTVMNSGALTSQVHTFPAVLQSSGLYSLSVTVSSNFGTQTYTCNVDHKPSNTKVDK 300
QY 208 ----KSCDKTHTCR-ELAGGSVFLEPPPKDITLMSRPEVTCVVDVSHEDPEVKRW 262
DB 301 TVERKCCVECPCPAPVAGPSVFLPPPKDITLMSRPEVTCVVDVSHEDPEVKRW 360
QY 263 YVDGVEVNAKTKPREBOYSTYRVVSVTLVHODMLNGEKYCKVSNKALPAPIEKTIS 322
DB 361 YVDGVEVNAKTKPREBOYSTYRVVSVTLVHODMLNGEKYCKVSNKALPAPIEKTIS 420
QY 323 KAKGQPREPOVYTLTPPSRDELTKNQVSLTCLVKGFYPSDIAVEMWSNGOPENNYKTTTPV 382
DB 421 KTKGQPREPOVYTLTPPSRDELTKNQVSLTCLVKGFYPSDIAVEMWSNGOPENNYKTTTPV 480
QY 383 LDSDSGFFLYSKLTVDKSRMOGQNFSCSVMEALHNHYTQKSLSLSPG 431
DB 481 LDSDSGFFLYSKLTVDKSRMOGQNFSCSVMEALHNHYTQKSLSLSPG 529

RESULT 26

AAVS1078 standard; protein; 729 AA.

XX AAVS1078;

XX 23-MAR-2000 (first entry)

XX Human fusion protein CD4H-1.

XX Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;

XX anti-human immunodeficiency virus; CD4H-1.

XX Homo sapiens.
OS Synthetic.
XX US6004781-A.

XX 21-DEC-1999.

XX 04-FEB-1994; 94US-00191708.

XX 22-JAN-1988; 88US-00147351.

XX 23-JAN-1989; 89US-00299596.

XX 09-JUN-1992; 92US-00896781.

XX 12-APR-1993; 93US-00057952.

XX (GENO) GEN HOSPITAL CORP.

XX Seed B;

XX MPI; 2000-085792/07.

XX N-F8DB; AAZ44061.

XX Fusion protein useful for the treatment of human immunodeficiency virus.

XX Example 1; Col 15-30; 39pp; English.

CC This invention describes a novel nucleic acid (I) encoding a fusion
CC protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC chain (III). The products of the invention have anti-human
CC immunodeficiency virus (HIV) activity and are capable of binding to
CC gp120. The fusion protein is useful for treating human immunodeficiency
CC virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC represents the fusion protein CD4H-1 which is constructed from CD4 linked
CC to human IgG1 upstream of the CH1 region

XX Sequence 729 AA;

Query Match 85.7%; Score 2069.5; DB 3; Length 729;
Best Local Similarity 58.7%; Pred. No. 9.4e-109;
Matches 427; Conservative 0; Mismatches 4; Indels 297; Gaps 4;

QY 1 MNRGVPFRLHLVQLALPAATQGNKVVLGKGGDTVELCTASQKSIQFMKNSNOIK 60
DB 1 MNRGVPFRLHLVQLALPAATQGNKVVLGKGGDTVELCTASQKSIQFMKNSNOIK 60
QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVQL 120
DB 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGILTANSDDTHLLOQGSITLTLESPPGSSBVOCSPRGKNIQGGKTLVSQLELDPSG 180
DB 121 LVFGILTANSDDTHLLOQGSITLTLESPPGSSBVOCSPRGKNIQGGKTLVSQLELDPSG 180
QY 181 TWTCTVLONOKKVEFKIDIV----- 200
DB 181 TWTCTVLONOKKVEFKIDIVLAAPASTKGPSPFLAPCSRSTSESTALGCLVKDYFPEP 240
QY 201 ----- 200
DB 241 QAERASSKSWITFDLKNKEVSVKRVYQDPKLGKGLPLHLTLPLQALPOYAGSGLTLA 300
QY 201 ----- 200
DB 301 LEAKTKLHDEVNLVVMKATOLQKNLTCEVWGPSTPKLMLSKLENKAKVSKREKPYWV 360
QY 201 -----PC----- 202
DB 361 LNPBAGMQLLSGQVLLSNIKVLPTWSTPVHADPEASTKGPSPFLAPSSKSTSGG 420
QY 203 -----PAP----- 205
DB 421 TAAAGCLVKDYFPEPVTVSMNSGALTSQVHTFPAVLQSSGLYSLSVTVTPSSSLGTQTY 480

```

QY 206 -----EPKSCDKTHTC-----PELLGGPSVFLFPPKPKDTLMISRTPE 243
DB 481 ICNVNKHPSNTFKVDKVEPKSCDKTHTCPCPCAPPELLGGPSVFLFPPKPKDTLMISRTPE 540
QY 244 VTCVVVDVSHEDPEVKFNNYV DGVGVNAKTKRREOQNSITRYRVSVLT VTHQOMLNKE 303
DB 541 VTCVVVDVSHEDPEVKFNNYV DGVGVNAKTKRREOQNSITRYRVSVLT VTHQOMLNKE 600
QY 304 YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 363
DB 601 YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 660
QY 364 VEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRMQQGNVFSCSVMHEALHNHYTQ 423
DB 661 VEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRMQQGNVFSCSVMHEALHNHYTQ 720
QY 424 KSLSLSPG 431
DB 721 KSLSLSPG 728

```

RESULT 27

AAP93012 standard; protein; 614 AA.

AAP93012;

25-MAR-2003 (revised)

03-AUG-1992 (first entry)

Genetic construct which encodes CD4 linked to human IgG1 at the BstI site downstream from the hinge region (fusion protein CD4BIambda).

Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy; diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.

Homo sapiens.

EP325262-A.

26-JUL-1989.

20-JAN-1989; 89EP-00100913.

22-JAN-1988; 88US-00147351.

(GENO) GEN HOSPITAL CORP.

Seed B;

WPI, 1989-214472/30.

N-PSDB; AAN90360.

Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV infections or detecting HIV or SIV in sample.

Example; Table 5, Page 48-55; 68pp; English.

The fusion protein genes of the invention pref. comprises cDNA sequences which encode CD4 or a fragment which binds gp120 ligated to an expression plasmid which encodes an antibody in which the variable region of the gene has been deleted (see WO87-02671). The CD4 portion of the fusion protein may comprise the complete CD4 sequence, the 370 AA extracellular region and the membrane spanning domain, or the extracellular region. The Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are specifically claimed: fusion proteins CD4BIambda1, CD4mu, CD4pnu, CD4BIambda1, and CD4mu (No. 67608), CD4BIambda (No. 67609) and CD4BIambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)

Sequence 614 AA;

Query Match 85.2%; Score 2056; DB 1; Length 614;

```

Best Local Similarity 67.6%; Pred. No. 4,5e-108;
Matches 416; Conservative 1; Mismatches 12; Indels 186; Gaps 4;
QY 1 MNRGVPFRHLLVLTQALLPATQGNKVVLGKGGDTVELCTASQKSIQFHMNSNQIK 60
DB 1 MNRGVPFRHLLVLTQALLPATQGNKVVLGKGGDTVELCTASQKSIQFHMNSNQIK 60
QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQNPFLIIKNLKI ESDTYICEVEDQKEEVOL 120
DB 61 ILNGSGFLTKGPSKLNDRADSRSLMDQNPFLIIKNLKI ESDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSTHLLQGQSLTLTLESPGSSPSVQCSPRKNIQGGKTLISVSOLELDSG 180
DB 121 LVFGLTANSTHLLQGQSLTLTLESPGSSPSVQCSPRKNIQGGKTLISVSOLELDSG 180
QY 181 TWICTVLQNGKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWICTVLQNGKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
QY 201 -----PKSCDKTHTC-----PCPAP----- 206
DB 241 QAEKSSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPLHLPLQALPYAGSGNLTLA 300
QY 207 -----PKSCDKTHTC----- 216
DB 301 LEAKTGLHGEVNLVVMRATOLQKNLTCEVWGPTSPMLSLKLENKAKVSRKEPVWV 360
QY 217 -----PELLGGPSVFLFPPKPKDTL 236
DB 361 LNPBAGMOCLLDSAOVLLESNIKVLPTWSTVHADPEAPPELLGGPSVFLFPPKPKDTL 420
QY 237 MISTPEVTCVVVDVSHEDPEVKFNNYV DGVGVNAKTKRREOQNSITRYRVSVLT VTHQ 296
DB 421 MISTPEVTCVVVDVSHEDPEVKFNNYV DGVGVNAKTKRREOQNSITRYRVSVLT VTHQ 480
QY 297 DWLNGKRYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVK 356
DB 481 DWLNGKRYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVK 538
QY 357 FYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRMQQGNVFSCSVMHEA 416
DB 539 FYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRMQQGNVFSCSVMHEA 598
QY 417 LHNHYTQKSLSPG 431
DB 599 LHNHYTQKSLSPG 613

```

RESULT 28

AAB19511 standard; protein; 616 AA.

AAB19511;

09-JAN-2001 (first entry)

CD4-IgG1 fusion protein CH4Bgammal.

CD4; IgG1; human; CD4Bgammal; fusion protein; immunoglobulin; HIV; SIV; gp120; therapy; diagnosis.

Homo sapiens.

Key location/Qualifiers

Protein 1..395 "CD4 extracellular region"

Protein 400..616 /note="IgG1 heavy chain"

US6117656-A.

12-SEP-2000.

PF 07-JUN-1995; 95US-00479353.
 XX
 PR 22-JAN-1988; 88US-00147351.
 PR 23-JAN-1989; 88US-00299596.
 PR 09-JUN-1992; 92US-0086781.
 PR 12-APR-1993; 93US-00057952.
 PR 04-FEB-1994; 94US-00191708.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Seed B;
 XX
 DR WPI; 2000-586558/55.
 DR N-PSDB; AAA50664.
 XX
 PT CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
 XX SIV.
 XX
 PS Example 1; Col 59-70; 39pp; English.
 XX
 CC The present sequence is that of fusion protein CD4gammal comprising the
 CC extracellular portion of CD4, which binds to HIV gp120, linked at its C-
 CC terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
 CC encoding CD4 was linked to IgG1 DNA at the BamI site downstream of the
 CC hinge region (see AAA50664). Fusion protein CD4gammal and a nucleic acid
 CC encoding it are claimed. Also claimed are a vector comprising the nucleic
 CC acid, and a method of producing the fusion protein in secreted form using
 CC a transformed host cell. The fusion protein may further comprise a
 CC therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
 CC can be administered to an animal (including humans) for treatment of HIV
 CC or SIV infection, and can also be used in assays for HIV or SIV, imaging
 CC and tissue stains. IgG1 fusion proteins such as CD4gammal provide both
 CC complement-mediated and cell-mediated immunity
 CC
 SQ Sequence 616 AA;
 Query Match 82.8%; Score 1999; DB 3; Length 616;
 Best Local Similarity 66.8%; Pred. No. 7, 5e-105;
 Matches 410; Conservative 1; Mismatches 19; Indels 184; Gaps 4;
 QY 1 NNRGVFRLHLVLTALIPATQGNKVVLGKKDVELTCTASQKSIQFMKNSNQIK 60
 DB 1 NNRGVFRLHLVLTALIPATQGNKVVLGKKDVELTCTASQKSIQFMKNSNQIK 60
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDGNFPLIINKIKEDSDTYICEVEDQKEVQL 120
 DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDGNFPLIINKIKEDSDTYICEVEDQKEVQL 120
 QY 121 LVFGLTANSDDLHLLQGSLLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 DB 121 LVFGLTANSDDLHLLQGSLLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
 QY 181 TWTCTVLQNKQKVEFKIDIV-----PCP----- 203
 DB 181 TWTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVFSFPLAFVTEKLTGSGELMW 240
 QY 204 -----APRPK----- 208
 DB 241 QAERASSKSWITFDLKNKEVSKRVTOPLQMGKGLPLHLTLPOALPOVAGSGLTLA 300
 QY 209 -----SCDKTHTC----- 216
 DB 301 LEAKTGKHLQEVNVLVVMRAVLQLOKNTLCVWGPTSPFKMLSLKLENKAKVSKREKPVW 360
 QY 217 -----PELGGPSVFLPPPKDNL 236
 DB 361 LNPAGMGQCLSDSGVLLLESNIKVLPTWSTPVHADPEAPILGSPVFLPPPKDNL 420
 QY 237 MTSRPEVYCVVVDVSHEDPEVKFNMYVUGVEVHNAKTPREBOYNSTTRVSVLTVLHQ 296
 DB 421 MTSRPEVYCVVVDVSHEDPEVKFNMYVUGVEVHNAKTPREBOYNSTTRVSVLTVLHQ 480
 QY 297 DWLNCKEYKCKVSNKALPAPIEKTISSKAGPREPOVYTLPPSRDELTKNGVSLTCLVKG 356

DB 481 DWLNCKEYKCKVSNKALPAPIEKTISSKAGPREPOVYTLPPSRDELTKNGVSLTCLVKG 540
 QY 357 FYPSDIAVWESWNSGCPENNYKTPPVLDSDSPFLYSLTUTDKSRMOGNFSCSVHMEA 416
 DB 541 FYPSDIAVWESWNSGCPENNYKTPPVLDSDSPFLYSLTUTDKSRMOGNFSCSVHMEA 600
 QY 417 LHNHYTOKSLSLSP 430
 DB 601 LHNHYTOKSLSLSP 614
 RESULT 29
 AAW35860
 ID AAW35860 standard; protein; 410 AA.
 XX
 AC AAW35860;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Human CD4:IgG2a constant region fusion protein.
 XX
 KW Human; CD4:IgG2a; T lymphocyte veto molecule; chimeric molecule;
 KW targeting polypeptide; suppression; immune response; treatment;
 KW autoimmune disease; allergy; immunological disorder; fusion protein;
 KW immunoglobulin G2a; transplant rejection; constant region.
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO9737687-A1.
 XX
 PD 16-OCT-1997.
 XX
 PF 10-APR-1997; 97WO-US005943.
 XX
 PR 10-APR-1996; 96US-00630172.
 XX
 PA (NABE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
 XX
 PI Staerz UD;
 XX
 DR WPI; 1997-512419/47.
 XX
 PT T lymphocyte veto molecule comprising response cell activating protein -
 PT linked to molecule that targets stimulator cell marker, used for
 PT selective suppression of immune response, e.g. prevention of graft
 PT rejection or treatment of auto-immune disease.
 PS Claim 40; Page 75-76; 309pp; English.
 XX
 CC A novel T lymphocyte veto molecule is a chimeric molecule comprising a
 CC protein, e.g. the present sequence, linked to a targeting polypeptide
 CC that binds a molecule, which differentiates a host cell from a tissue
 CC graft cell, or selectively targets a stimulator cell involved in the
 CC autoimmune response. A veto molecule, in which the protein binds a
 CC molecule that targets stimulator cells, can be used to suppress an immune
 CC response and therefore treat autoimmune diseases, e.g. systemic lupus
 CC erythematosus, myasthenia gravis, rheumatoid arthritis, insulin dependent
 CC diabetes mellitus, multiple sclerosis, coeliac disease, autoimmune
 CC thyroiditis, Addison's or Grave's diseases and rheumatoid carditis,
 CC allergies and other immunological disorders. Where the protein binds a
 CC molecule that differentiates graft and host cells, the veto molecule can
 CC be used to reduce transplant rejection. The veto molecule provides
 CC specific regulation of particular stimulator cells that can kill graft
 CC cells or respond to autoantigens, but leave other stimulator cells
 CC unaffected, e.g. CD4 or CD8 positive cells can be regulated without one
 CC affecting the other. The veto molecule can be administered locally to
 CC minimise generalised immunosuppression
 CC
 SQ Sequence 410 AA;
 Query Match 68.2%; Score 1647.5; DB 2; Length 410;

Best Local Similarity 75.8%; Pred. No. 3.3e-85;
Matches 316; Conservative 31; Mismatches 51; Indels 19; Gaps 4;

```

QY 26 NKVVLGKKGDPVELTCTASQKSIQPFHMKNSNOIKILGNQGSFLTKGSKLNDADSRSS 85
DB 1 NKVVLGKKGDPVELTCTASQKSIQPFHMKNSNOIKILGNQGSFLTKGSKLNDADSRSS 60
QY 86 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLOQGSLLTLLES 145
DB 61 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLOQGSLLTLLES 120
QY 146 PPGSSPSVQCRSPRKNIQGGKTLVSQLELDQSGTWTCTVLONOKKVEFKIDIV---- 200
DB 121 PPGSSPSVQCRSPRKNIQGGKTLVSQLELDQSGTWTCTVLONOKKVEFKIDIVLVAEP 180
QY 201 -----PCPAPRPSKCDKHTHCPELLGGPSVFLPFPKPDITLMSRTPEVLCVVAVDSHE 254
DB 181 RGPRTKPCP---PCNC-----PAPNLGGPSVFLPFPKIDVLMSTLSPTVCVVAVDSSED 233
QY 255 DPEVFNWYVDGVEVHNNAKTKPREQYNSTYRVVSVLTVLHODWLNGKEYKCKVSNKALP 314
DB 234 DPDVQISFVNNVVEHTAQOTGTHREDYNSRLRVVSALPILQGDMMWSGKEFKCKVNNKDLP 293
QY 315 APIETTKSKAQPRPEPOVYTLPRSRDELTKNQVSLTCLVKGFIPSDIAVWESNGOPEN 374
DB 294 APIERTIKRPGSVAPQVYVLPFP-EEMTKKQVTLTCVTDTFMPEEDITYEMVTNGKTEL 352
QY 375 NYKTPPVLDSGSPFLSKLTVDSRMQOGNVSQVMHEALNHYQKSLSPG 431
DB 353 NYKTEPVLDSGSPFLSKLTVDSRMQOGNVSQVMHEALNHYQKSLSPG 409

```

```

RESULT 30
AAR20634
ID AAR20634 standard; protein; 343 AA.
XX
AC AAR20634;
XX
DT 22-MAY-1992 (first entry)
XX
DE Chimeric protein V1V2-hCH2-KA.
XX
KM CD4, LTI; Streptomyces longisporous; HIV gp120; AIDS; IgG1; T cell;
KW immunoglobulin; surface glycoprotein; virus; MHC class II.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cleavage-site 28..29
FT Protein /label= signal_cleavage
FT Region /label= V1V2
FT Protein /label= hinge
FT Protein /label= CH2
XX
XX MO3200985-A.
XX
XX 23-JAN-1992.
XX
XX 11-JUL-1990; 90US-00551584.
XX
XX 11-JUL-1990; 90US-00551584.
XX
XX 05-MAR-1991; 91US-00665218.
XX
XX (SMIK ) SMITHKLINE BEECHAM.
XX
XX Brawner ME, Fornwald JA, Arthos J;
XX
XX WPI; 1992-056814/07.
XX
XX DR N-PSDB; AAQ20907.
XX

```

PT Nucleic acid sequences for production of CD4 chimeric protein - used to
PT transfect streptomyces, contg. LTI signal sequence linked to pro-peptide
PT sequence facilitating peptide cleavage.

PS Claim 16; Page 23; 47pp; English.

XX The sequence was deduced by sequencing the plasmid vector V1V2-hCH2-KA
XX in S. lividans strain 1326. The protein has domains contg. peptides of
XX different function. It contains a CD4 chimera (V1V2) in which the carboxy
XX terminal portion of the protein consists of a murine immunoglobulin light
XX chain constant region, linked to the signal peptide of Streptomyces LTI,
XX modified at its N-terminus to include Lys-Arg. Also included is the IgG1
XX constant region comprising the hinge and CH2 motifs. Human IgG1 is the
XX most effective immunoglobulin subclass at mediating cell killing by both
XX complement and ADCC. The CD4 chimeric proteins may be expressed in
XX bacterial hosts. The fusion of the human Ig constant region lacking the
XX CH3 domain, and the LTI CD4 protein increases the stability of the CD4,
XX thus increasing the serum half life and/or potency against HIV infection
XX and inhibit virus-induced cell fusion, relative to soluble CD4. By
XX altering only one amino acid at position 2 near the N terminal of CD4 (V1
XX region) from Lys to Ala, a heterologous protein is expressed which is
XX efficiently secreted and correctly processed to remove the entire LTI
XX signal sequence, but which still retains the gp120 binding capacity. By
XX modifying the pro-peptide you avoid deleterious effects of additional
XX amino acids on the function of the protein. See also AAR20635.6

SO Sequence 343 AA;

Query Match 64.0%; Score 1546; DB 2; Length 343;
Best Local Similarity 93.2%; Pred. No. 1.5e-79;
Matches 300; Conservative 0; Mismatches 12; Indels 10; Gaps 2;

```

QY 19 LPATQGNKVVVLGKKGDPVELTCTASQKSIQPFHMKNSNOIKILGNQGSFLTKGSKLND 78
DB 22 LTGAALAKAVVLGKKGDPVELTCTASQKSIQPFHMKNSNOIKILGNQGSFLTKGSKLND 81
QY 79 RADSRRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLOQGS 138
DB 82 RADSRRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLOQGS 141
QY 139 LTTLTBSPSSPSVQCRSPRKNIQGGKTLVSQLELDQSGTWTCTVLONOKKVEFKID 198
DB 142 LTTLTBSPSSPSVQCRSPRKNIQGGKTLVSQLELDQSGTWTCTVLONOKKVEFKID 201
QY 199 IV-----PCPAPRPSKCDKHTHC-----PELLGGPSVFLPFPKPDITLMSRTPEVLCV 248
DB 202 IVVLAPQKASKVPEKSCDKHTHCPCPAPRPLGGPSVFLPFPKPDITLMSRTPEVLCV 261
QY 249 VDVSHEDEPEVKFNWYVDGVEVHNNAKTKPREQYNSTYRVVSVLTVLHODWLNGKEYKCKV 308
DB 262 VDVSHEDEPEVKFNWYVDGVEVHNNAKTKPREQYNSTYRVVSVLTVLHODWLNGKEYKCKV 321
QY 309 SNKALPAPIETKTSKAKGPRE 330
DB 322 SNKALPAPIETKTSKAKGPRE 343

```

```

RESULT 31
AAR89441
ID AAR89441 standard; peptide; 254 AA.
XX
AC AAR89441;
XX
DT 26-SEP-1996 (first entry)
XX
XX IgG1 hinge, CH2 and CH3 domains.
XX
XX CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
XX human; CD4; HIV; proteinoaceous alpha-helix; T cell; B cell; neutrophil;
XX dendritic cell; therapy; mammal; infection.
XX
XX Homo sapiens.
XX

```

```
PN W09603883-A1.
XX
XX 15-FEB-1996.
XX
XX 26-JUL-1995; 95WO-US009468.
XX
XX 02-AUG-1994; 94US-00284391.
PR 24-FEB-1995; 95US-00394388.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Seed B, Banapour B, Romeo C, Kolanus W;
PI WPI; 1996-129034/13.
DR N-PSDB; AAT10780.
XX
XX Membrane-bound chimeric receptor comprising extracellular portion
PT including CD4 fragment - cells expressing receptor can be used for
PT treatment of HIV infection.
XX
XX Claim 3; Fig 25; 134pp; English.
XX
XX This sequence represents the human IgG1 hinge, CH2 and CH3 domains. This
CC sequence is included in the membrane bound proteinaceous chimeric
CC receptor of the invention. Alternatively the transmembrane region of the
CC chimeric receptor contains a portion of the CD7, CD5 or CD34
CC transmembrane domains. The extracellular portion of the chimeric receptor
CC contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
CC sequence) which specifically recognizes and binds HIV-infected cells, but
CC does not mediate HIV infection. The extracellular domain of the receptor
CC is separated from the cell membrane by 48 or 72 angstroms, or by one or
CC more proteinaceous alpha-helices. The cells expressing the receptor are
CC preferably T cells, B cells, neutrophils, or dendritic cells. The
CC therapeutic cells expressing the chimeric receptor are administered to a
CC mammal to treat HIV infection
XX
XX Sequence 254 AA;
SQ
Query Match 55.4%; Score 1338.5; DB 2; Length 254;
Best Local Similarity 98.0%; Pred. No. 5.5e-68;
Matches 249; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
QY 206 EPKSCDKHTHTC-----PELLGGPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKF 260
DB 1 EPKSCDKHTHTCPPCAPPELLGGPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKF 60
QY 261 NMYVDGVEVHNAKTKRREQDYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT 320
DB 61 NMYVDGVEVHNAKTKRREQDYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT 120
QY 321 ISKAKGQPREPOVYTLPPSRDELITKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTTP 380
DB 121 ISKAKGQPREPOVYTLPPSRDELITKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTTP 180
QY 381 PVLDSGSEFLLYSKLTVDKSRMQQGNVSCVMHEALHNYTKQSLSLSPGLQDETCAE 440
DB 181 PVLDSGSEFLLYSKLTVDKSRMQQGNVSCVMHEALHNYTKQSLSLSPGLQDETCAE 240
QY 441 AODGELDGLMTTDP 454
DB 241 AODGELDGLMTTDP 254
RESULT 32
AAR78667
ID AAR78667 standard; protein; 254 AA.
XX
XX AAR78667;
AC
XX
XX 11-APR-1996 (first entry)
DT
XX
XX IgG1 hinge, CH2 and CH3 domains.
DE
XX
```

```
KW Chimeric receptor; CD4; T-cell receptor; HIV; cytolysis;
KW human immunodeficiency virus; adoptive immunotherapy; IgG1.
XX
XX Homo sapiens.
XX
XX W09521528-A1.
XX
XX 17-AUG-1995.
XX
XX 12-JAN-1995; 95WO-US000454.
XX
XX 14-FEB-1994; 94US-00195395.
PR 02-AUG-1994; 94US-00284391.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Seed B, Banapour B, Romeo C, Kolanus W;
PI WPI; 1995-292893/38.
DR N-PSDB; AAQ96101.
XX
XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
PT cells.
XX
XX Claim 3; Fig 25; 118pp; English.
XX
XX Human IgG1 hinge, CH2 and CH3 domains (AAR78668) are used in the
CC construction of a chimeric receptor utilised in the targeted cytolysis of
CC HIV-infected cells. The chimeric receptor comprises the extracellular
CC domain (pref. amino acids 1-394 or 1-200) of CD4 linked via the CD7
CC transmembrane domain to an intracellular portion, e.g. of T-cell receptor
CC protein zeta. The IgG1 portion of the chimeric receptor is encoded by the
CC DNA sequence given in AAQ96101
XX
XX Sequence 254 AA;
SQ
Query Match 55.2%; Score 1332.5; DB 2; Length 254;
Best Local Similarity 97.6%; Pred. No. 1.2e-67;
Matches 248; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
QY 206 EPKSCDKHTHTC-----PELLGGPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKF 260
DB 1 EPKSCDKHTHTCPPCAPPELLGGPSVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKF 60
QY 261 NMYVDGVEVHNAKTKRREQDYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT 320
DB 61 NMYVDGVEVHNAKTKRREQDYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT 120
QY 321 ISKAKGQPREPOVYTLPPSRDELITKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTTP 380
DB 121 ISKAKGQPREPOVYTLPPSRDELITKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTTP 180
QY 381 PVLDSGSEFLLYSKLTVDKSRMQQGNVSCVMHEALHNYTKQSLSLSPGLQDETCAE 440
DB 181 PVLDSGSEFLLYSKLTVDKSRMQQGNVSCVMHEALHNYTKQSLSLSPGLQDETCAE 240
QY 441 AODGELDGLMTTDP 454
DB 241 AODGELDGLMTTDP 254
RESULT 33
ADD13790
ID ADD13790 standard; protein; 400 AA.
XX
XX ADD13790;
AC
XX
XX 01-JAN-2004 (first entry)
DT
XX
XX Plasmid pBS loxp-IgG1/pBS loxp-IgG1delta350/pBS loxpIggdeltaCHI protein.
DE library; transfection; humanized monoclonal antibody; antigen;
KW T cell receptor; circular.
KW
```

```

XX OS Synthetic.
OS Homo sapiens.
OS Mus sp.
XX FH Key
XX FT Location/Qualifiers
XX FT 1..97
XX FT /note="Human IgG1 CH3"
XX FT Region
XX FT 98..112
XX FT /note="Human IgG1 hinge"
XX FT Region
XX FT 113..222
XX FT /note="Human IgG1 CH2"
XX FT Region
XX FT 223..329
XX FT /note="Human IgG1 CH3"
XX FT Region
XX FT 330..373
XX FT /note="Murine IgG1 M1"
XX FT 374..400
XX FT /note="Murine IgG2 M2"
XX PN EP1298207-A1.
XX PD 02-APR-2003.
XX PF 01-OCT-2001; 2001EP-00123596.
XX PR 01-OCT-2001; 2001EP-00123596.
XX PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX PI Breitling F, Moldenhauer G, Poustka A, Kuehlwein T;
XX DR WPI, 2003-383833/37.
XX PT Preparing library of protein-producing eukaryotic cells, useful for
XX PT producing humanized high-affinity antibodies, comprises introducing
XX PT specific recombination signals into chromosomal gene loci and integrating
XX PT a variety of DNA sequences.
XX PS Example 19; Fig 16; 75pp; German.
XX CC This invention describes a novel method of preparing a library of protein
XX CC -producing eukaryotic cells comprising (a) introducing specific
XX CC recombination signals into one or two chromosomal gene loci, (b)
XX CC Expanding at least one of the modified cells, (c) Transfecting many
XX CC different DNA sequences, each flanked by recombination signals, into the
XX CC expanded cells and (d) Integrating the DNA sequences into the gene loci
XX CC on the basis of the recombination signals and the appropriate
XX CC recombinase. The resulting cells express different proteins, each from an
XX CC integrated DNA sequence and the proteins are bound to the cell surface.
XX CC The method is particularly used to produce libraries of humanized
XX CC monoclonal antibodies, for selection of those with affinity for
XX CC particular antigens and useful for diagnostic or therapeutic use.
XX CC Libraries of T cell receptors may also be prepared. The method produces
XX CC libraries of high diversity; provides easy, quick and automatable
XX CC selection from a large number of proteins, allows relatively simple
XX CC alteration of the expressed gene (e.g. fusion to other protein-coding
XX CC sequences), is suitable for large scale protein production and allows
XX CC simple verification and characterization of selected cell lines. The
XX CC method does not require incorporation of a resistance marker. This
XX CC sequence represents the construct pBS loxp-IgG1/pBS loxp-IgG1delta50/pBS
XX CC loxpIgdeltaCh1 described in the disclosure of the invention.
XX SQ Sequence 400 AA;
XX
Query Match 55.0%; Score 1327.5; DB 7; Length 400;
Best Local Similarity 75.4%; Pred. No. 3.7e-67;
Matches 266; Conservative 8; Mismatches 22; Indels 57; Gaps 7;
Oy 151 PSVQCRSPRGKNIQGG-----KTLVS-----QLEIQDSG-----180
Db 5 PSVFLAASSKSTSGTALGLVDVDPPEPTVSWNSGALTSVHTPPAVLQSSGLYSL 64
Oy 181 -----TWCTVLQNRKVEFKIDIVPCAPPEPKSCDKHTTC-----PELLG 221

```

```

Db 65 SSVTVPSSLGTYICNV--NHKSNTRYD---KKVPEKSCDKHTCPPCAPPELLG 118
Oy 222 GPSVFLPPEKPKOTLMISRTPEVTCVVVDVSHEDPEVKFMWYDGVVHNAKTRPREQY 281
Db 119 GPSVFLPPEKPKOTLMISRTPEVTCVVVDVSHEDPEVKFMWYDGVVHNAKTRPREQY 178
Oy 282 NSTYRVSVLTVLHODWLNGEKYCKVSNKALPAPIEKITSKAGQPREPOVYTLPPSRD 341
Db 179 NSTYRVSVLTVLHODWLNGEKYCKVSNKALPAPIEKITSKAGQPREPOVYTLPPSRD 238
Oy 342 ELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 401
Db 239 ELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 298
Oy 402 WQGNVPSGCVMEHAIHNYTOKSLSP--GLQDETCAEADQGEIDGIMTT 452
Db 299 WQGNVPSGCVMEHAIHNYTOKSLSP--GLQDETCAEADQGEIDGIMTT 351

RESULT 34
ADD13781
ID ADD13781 standard; protein; 401 AA.
XX AC ADD13781;
XX DT 01-JAN-2004 (first entry)
XX DE Plasmid pBS MbigIM/ pBS MbigIMdelta250 protein.
XX KW library; transfection; humanized monoclonal antibody; antigen;
XX KW T cell receptor; circular.
XX OS Synthetic.
XX OS Homo sapiens.
XX OS Mus sp.
XX FH Key
XX FT Location/Qualifiers
XX FT 1..97
XX FT /note="human IgG1 CH1"
XX FT Region
XX FT 98..112
XX FT /note="human IgG1 hinge"
XX FT Region
XX FT 113..222
XX FT /note="human IgG1 CH2"
XX FT Region
XX FT 223..330
XX FT /note="human IgG1 CH3"
XX FT Region
XX FT 331..374
XX FT /note="murine IgG1 M1"
XX FT 375..401
XX FT /note="murine IgG1 M2"
XX PN EP1298207-A1.
XX PD 02-APR-2003.
XX PF 01-OCT-2001; 2001EP-00123596.
XX PR 01-OCT-2001; 2001EP-00123596.
XX PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX PI Breitling F, Moldenhauer G, Poustka A, Kuehlwein T;
XX DR WPI, 2003-383833/37.
XX DR N-PSDB; ADD13780.
XX PT Preparing library of protein-producing eukaryotic cells, useful for
XX PT producing humanized high-affinity antibodies, comprises introducing
XX PT specific recombination signals into chromosomal gene loci and integrating
XX PT a variety of DNA sequences.
XX PS Example 1; Fig 12B; 75pp; German.

```

CC This invention describes a novel method of preparing a library of protein
CC -producing eukaryotic cells comprising (a) introducing specific
CC recombination signals into one or two chromosomal gene loci; (b)
CC Expanding at least one of the modified cells; (c) Transfecting many
CC different DNA sequences, each flanked by recombination signals, into the
CC expanded cells and (d) Integrating the DNA sequences into the gene loci
CC on the basis of the recombination signals and the appropriate
CC recombinase. The resulting cells express different proteins, each from an
CC integrated DNA sequence and the proteins are bound to the cell surface.
CC The method is particularly used to produce libraries of humanized
CC monoclonal antibodies, for selection of those with affinity for
CC particular antigens and useful for diagnostic or therapeutic use.
CC Libraries of T cell receptors may also be prepared. The method produces
CC libraries of high diversity; provides easy, quick and automatable
CC selection from a large number of proteins, allows relatively simple
CC alteration of the expressed gene (e.g. fusion to other protein-coding
CC sequences), is suitable for large scale protein production and allows
CC simple verification and characterization of selected cell lines. The
CC method does not require incorporation of a resistance marker. This
CC sequence represents the construct Mb1g1M/ PBS Mb1g1Mdelta250 described
CC in the disclosure of the invention.

SQ Sequence 401 AA;

Query Match 54.5%; Score 1315; DB 7; Length 401;

Best Local Similarity 75.1%; Pred. No. 1,9e-66;

Matches 266; Conservative 7; Mismatches 23; Indels 58; Gaps 8;

QY 151 PSVQCRSPRKXNIQGG-----KTLVS-----QLELDQSG---- 180

DB 5 PSVFLPAPASKSTSGTALGCLVXDYPPEVTVSNVSGALTSVHTFPAYLQSSGLYSL 64

QY 181 -----TWTCTVLONQKVEBKIDIVPAPAEPSCDTHTC-----PELLG 221

DB 65 SSVVTVPSSLSGTQYTCNV--NHRKPSNTKVD---KXVEPKSCDTHTCPPCAPPELLG 118

QY 222 GPSVFLPAPKPDITLMISRTPEVTCVVDVSHEDPEKFNMYVDGVEVNAKTKREEDY 281

DB 119 GSVFLPAPKPDITLMISRTPEVTCVVDVSHEDPEKFNMYVDGVEVNAKTKREEDY 178

QY 282 NSTYRVVSVLTVLHODMNGKEYCKVSNKALPAPIEKTISRAGQPREPOVYTLPPSRD 341

DB 179 NSTYRVVSVLTVLHODMNGKEYCKVSNKALPAPXEKTISRAGQPREPOVYTLPPSRD 238

QY 342 ELTKNOVSLTCLVKGYPSPDIWEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSR 401

DB 239 ELTKNOVSLTCLVKGYPSPDIWEMESNGOPENNYKTPPVLDSDGSFFLYSKLTVDKSR 298

QY 402 WQGNVFCSCVMHEALHNHYTKSL-SLSP--GLQDETCAEADQDELGLMTT 452

DB 299 WQGNVFCSCVMHEALHNHYTKSSLSLSPGQLDDETCARQDELGLMTT 352

RESULT 35

ABJ37104

ID ABJ37104 standard; protein; 437 AA.

XX ABJ37104;

XX 08-MAY-2003 (first entry)

DE Concatameric immunoadhesion human protein sequence SEQ ID No 14.

XX Antinflammatory; antibacterial; immunosuppressive; antineumatic;

XX antiarthritic; immunomodulator; concatameric protein; soluble domain;

XX dimeric protein; inflammation; septicemia; cytotoxicity;

XX rheumatoid arthritis; cachexia; inflammation; human.

OS Homo sapiens.

XX WO2003010202-A1.

XX 06-FEB-2003.

XX 26-JUL-2002; 2002WO-KR001427.

XX 26-JUL-2001; 2001KR-00045028.

XX (MEDE-) MEDEXGEN CO LTD.

XX Chung Y, Han J, Lee H, Choi E, Kim J;

XX WPI, 2003-229639/22.

XX N-PSDB; AB732047.

XX New concatameric protein having two soluble domains, useful for

XX diagnosing and treating disorders associated with the dimeric protein or

XX its glycosylated form, such as inflammation, septicemia, rheumatoid

XX arthritis and cachexia.

XX Disclosure; Page 156-158; 21pp; English.

XX The invention relates to a novel concatameric protein comprising two

XX soluble domains, in which an N-terminus of a soluble domain of a

XX biologically active protein is linked to a C-terminus of an identical

XX soluble domain or a different soluble domain of a biologically active

XX protein. The methods and compositions of the present invention are useful

XX for the diagnosis and treatment of disorders associated with dimeric

XX protein or its glycosylated form, such as inflammation, septicemia,

XX cytotoxicity, rheumatoid arthritis, cachexia and other inflammatory

XX related diseases. This sequence represents the human concatameric protein

XX of the invention

SQ Sequence 437 AA;

Query Match 53.3%; Score 1287.5; DB 6; Length 437;

Best Local Similarity 64.7%; Pred. No. 7.3e-65;

Matches 273; Conservative 22; Mismatches 62; Indels 65; Gaps 11;

QY 35 DTVELTCTAASOKSIOFMKNSNQIKIUNGQSFYTKPSKLNDRASRLMDQGNPPL 94

DB 55 DDIKEKTSDDKKAQAFKEKE-----TFREKTYTLFK-----NGTL 92

QY 95 IINKLIKEDSDTYICEVEDQK-EVYQLLVFGLTANSDTHLQOQ-----SLTLYT 142

DB 93 KIKHLKTDODIYKVSIVDTKGKVLKIFDLK-----IQERVSKPISWTCINTLYT 145

QY 143 LESPPGSPSVQCRSPRKXNIQGGKTLVSQLELDQSGTWT-----CTVLONQKVE 194

DB 146 CEVWNGTDPENLNL-----YQDGRHLKLSQRYI--THKWTTSLSAKFKCTA-GNKYSKE 195

QY 195 FKIDIVPAPAEPSKCDTHTC-----PELLGSPVFLPAPKPDITLMISRTPEVTCVV 249

DB 196 SSVFVSGCPA-EPPSCDTHTCPPCAPPELLGGSVFLPAPKPDITLMISRTPEVTCVV 254

QY 250 DVSHEDPEKFNMYVDGVEVNAKTKPREEQNSTYRVVSVLTVLHODMNGKEYCKVCS 309

DB 255 DVSHEDPEKFNMYVDGVEVNAKTKPREEQNSTYRVVSVLTVLHODMNGKEYCKVCS 314

QY 310 NKALPAPIEKTISRAGQPREPOVYTLPPSRDELTKNOVSLTCLVKGYPSPDIWEMESN 369

DB 315 NKALPAPIEKTISRAGQPREPOVYTLPPSRDELTKNOVSLTCLVKGYPSPDIWEMESN 374

QY 370 GOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOGNVFCSCVMHEALHNHYTKSSL 429

DB 375 GOPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOGNVFCSCVMHEALHNHYTKSSL 434

RESULT 36

AAR42162

ID AAR42162 standard; protein; 461 AA.

XX

XX

```

AC AAR42162;
XX
DT 25-MAR-2003 (revised)
DT 27-APR-1994 (first entry)
XX
DE Anti-HIV-1 recombinant antibody 447-52D heavy chain.
XX
KW Human Immunodeficiency Virus; antigen; ELISA; recombinant antibody;
KW HIV-neutralising monoclonal antibody; immunoglobulin; AIDS;
KW acquired immune deficiency syndrome; chimeric antibody;
KW surface glycoprotein gp120; V3 loop.
XX
OS Homo sapiens.
XX
PN MO9319785-A1.
XX
PD 14-OCT-1993.
XX
PF 23-MAR-1993; 93WO-US002629.
XX
PR 01-APR-1992; 92US-00861701.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Emimi EA, Conley AJ, Mark GE, Johnson LS, Pfarr DS;
XX
DR WPI, 1993-336600/42.
XX
DR N-PSDB; AA049834.
XX
PT New recombinant human antibody - with HIV neutralising activity against
PT at least two isolates, useful for preventing or treating infection in
PT diagnosis, etc.
XX
PS Example 9; Fig 2A; 154pp; English.
XX
CC EBV-transformed cell lines and mouse-human heterohybridomas producing
CC human MAb specific for the gp120 V3 loop of HIV-1 MN isolate were
CC obtained. MAb 447-52D was found to recognise the tetrapeptide motif GPGR,
CC i.e. the Principal Neutralising Determinant common to the V3 loop of
CC different HIV isolates. A recombinant Ab was produced in which the H
CC chain V region was derived from 447-52D and to which a signal sequence
CC and a H chain intronic sequence are appended, fused to a fragment contg.
CC a short intronic segment of the human gamma 1 C region and the human
CC gamma 1 encoding domain in its genomic form. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX
SQ Sequence 461 AA;

Query Match 53.3%; Score 1287; DB 2; Length 461;
Best Local Similarity 59.1%; Pred. No. 8.2e-65;
Matches 276; Conservative 30; Mismatches 79; Indels 82; Gaps 11;

QY 30 LGKGDVVELTCTAS--QKSIQFHWKNSNQIKILGNQGSFLTKGPSKLN----- 78
DB 11 LKPGGSLRLTCVASGFTSPDVMWVQAPGKLEWGRIRKSDGGTTVAASVKGRF 70
QY 79 ---RADSRSLMDQGNPLIKNLKIEDSDTYICEVEDQKEEVLVFGLTANSDTHLQ 135
DB 71 TISRSDSKNTLYLQMN-----SLKTEPTAVYSCTTDG-----FLMIGVEDYIYYMD 119
QY 136 --GOSLTLTLESPGSSPSVOCSPRGKNIQGG-----KTLVS----- 172
DB 120 VMGKQTTVTSASATKGPSVFPPLAPSSKTSQGTALGCLVKDYFPEPVTVSWMNSGALT 179
QY 173 -----QLELDQSG-----TWCTVYQONQKYEKFDIYPCAPPEPKS 209
DB 180 GVHTFPAVLQSSGLYSLSVTVPSSSLGTQTYICNV--NHKPSNTKVD-----KKEPEKS 233
QY 210 CDKHTHTC-----PELLGGPSVFLFPKPKDITLMSRTEPVQCVVDVSHEDPEVKPMYV 264
DB 234 CDKHTHTCPCAPFELGGLGPSVFLFPKPKDITLMSRTEPVQCVVDVSHEDPEVKPMYV 293
QY 265 DGEVHNNAKTPREEQNSTYRVVSVLTVLHQDWLNGEKYCKVSNKALPAPIEKITSKA 324

```

```

DB 294 DGEVHNNAKTPREEQNSTYRVVSVLTVLHQDWLNGEKYCKVSNKALPAPIEKITSKA 353
QY 325 KGPPEPQVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAVEMWESNGQPENNYKTPPVLD 384
DB 354 KGPPEPQVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAVEMWESNGQPENNYKTPPVLD 413
QY 385 SDGSFFLYSKLTVDKSRMOQGNFSCSVMEALHNHYTKQSLSPG 431
DB 414 SDGSFFLYSKLTVDKSRMOQGNFSCSVMEALHNHYTKQSLSPG 460

RESULT 37
AA018400
ID AA018400 standard; protein; 449 AA.
XX
AC AA018400;
XX
DT 11-OCT-2002 (first entry)
XX
DE Mature humanised murine CBE11 heavy chain variable domain.
XX
KW Mouse; humanised antibody; lymphotoxin beta receptor; antibody; cancer;
KW neoplasia; LT-beta-R; light chain; heavy chain; variable region.
XX
OS Mus sp.
XX
OS Synthetic.
XX
PN W0200230986-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US032140.
XX
PR 13-OCT-2000; 2000US-0240285P.
PR 13-MAR-2001; 2001US-0275289P.
PR 21-JUN-2001; 2001US-0299987P.
XX
PA (BIOJ ) BIOGEN INC.
XX
PI Garber E, Lyne P, Saldanha JW;
XX
DR WPI; 2002-583337/62.
XX
PT New humanized anti-lymphotoxin-beta receptor antibody, useful for
PT treating or reducing the advancement, severity or effects of neoplasia,
PT particularly solid tumors (i.e. carcinomas) including colorectal cancer
PT and breast cancer.
XX
PS Example 5; Page 25-26; 41pp; English.
XX
CC The present invention relates to humanised anti-lymphotoxin beta receptor
CC (LT-beta-R) antibodies. These are derived from the murine LT-beta-R
CC binding antibody CBE11 and can be used to treat neoplasia in humans. The
CC present sequence is a humanised murine CBE11 heavy chain variable region
XX
SQ Sequence 449 AA;

Query Match 53.3%; Score 1285.5; DB 5; Length 449;
Best Local Similarity 59.0%; Pred. No. 9.7e-65;
Matches 275; Conservative 29; Mismatches 71; Indels 91; Gaps 13;

QY 30 LGKGDVVELTCTAS--QKSIQFHW-----KNSNQIKILGNQGSFLTKGPSKLN----- 77
DB 11 LKPGGSLRLSCASGFTSPDYIMYFROAPGKLEWATISDGSY--TYPPSVKGRFT 69
QY 78 -DRASRSLMDQGNPLIKNLKIEDSDTYICEVEDQKEEVLVFGLTANSDTHLQ- 135
DB 70 ISRDNAKNSLY-----LQMSLRAEDTAVYCARAE-----NGNFFYFDY 109
QY 136 --GOSLTLTLESPGSSPSVOCSPRGKNIQGG-----KTLVS----- 172
DB 110 WGGQTTVTVSASATKGPSVFPPLAPSSKTSQGTALGCLVKDYFPEPVTVSWMNSGALTSG 169

```